ENGINEERED TERMINAL NUCLEOTIDYL TRANSFERASE VARIANTS AND USES THEREOF

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application 63/634,882, filed April 16, 2024, and U.S. Provisional Application , filed 63/688,785, filed August 29, 2024, the contents of which are incorporated by reference herein.

REFERENCE TO SEQUENCE LISTING, TABLE OR COMPUTER PROGRAM

[0002] The official copy of the Sequence Listing is submitted concurrently with the specification as an XML file, with a file name of "CX10-268WO3_ST26.xml", a creation date of April 16, 2025, and a size of 28,175 bytes. The Sequence Listing filed is part of the specification and is incorporated in its entirety by reference herein.

TECHNICAL FIELD

[0003] The present disclosure provides engineered terminal nucleotidyl transferase (TnT) polypeptides useful in template-independent polynucleotide synthesis, as well as compositions and methods of utilizing these engineered polypeptides.

BACKGROUND

[0004] Synthetic biology is becoming established in a diverse range of high value, high growth markets. From food and agriculture to therapeutics, diagnostics, and vaccines; tools such as gene editing, DNA sequencing and gene synthesis are being used to build value-added products with advanced functionality (e.g., cell bioreactors, etc.) and desired end products (e.g., drugs, chemicals, etc.). The barrier to widespread implementation of these technologies is the ability to efficiently synthesize RNA, DNA, and other polynucleotides.

[0005] In particular, silencing RNA (siRNA) therapeutics are a promising class of drugs that have the potential to treat numerous difficult to treat conditions in a highly targeted manner by binding to known mRNA targets (Hu et al. (2020). Sig Transduct Target Ther 5, 101; Zhang et al. (2021). Bioch. Pharmac., 189, 114432.) As these therapies become more common and are targeted at larger patient populations, the ability to produce large amounts of the oligonucleotide active pharmaceutical ingredient (API) becomes critical.

[0006] To date, short RNA oligonucleotides have been synthesized almost exclusively by iterative addition of nucleotides in the form of activated phosphoramidites, plus additional processing steps, to a growing immobilized nucleotide chain (Brown, T. Nucleic Acids Book. See at: www.atdbio.com/nucleic-acids-book (accessed 2022-10-10).)

[0007] Phosphoramidite chemistry has been developed extensively over the years to synthesize small amounts of DNA and for more complex therapeutic RNA syntheses but suffers from several cost and sustainability issues that are potentially limiting as API demand grows to triple-quadruple digit kilograms per year (Andrews et al. (2021). J. Org. Chem. 86, 49–61). Additionally, RNA synthesis using phosphoramidite synthesis chemistry is limited to producing short oligonucleotides of approximately 200 basepairs (Beaucage & Caruthers. (1981). Tetrahedron Lett. 22 (20): 1859.)

[0008] The phosphoramidite iterative methodology is multi-step and based on phosphorous (III) coupling chemistry that requires (i) coupling (ii) capping (iii) oxidation to P(V) forming phosphodiester or phosphorothioate diester (iv) deblocking of 5'O group. After chain synthesis is complete, the final oligo is cleaved from the support where deblocking of phosphate cyanoethyl group and nucleobases can also occur (Brown, T. Nucleic Acids Book. See at: www.atdbio.com/nucleic-acids-book (accessed 2022-10-10).) Washes with organic solvents at each step are also required. The phosphate cyanoethyl blocking group and nucleobase protecting groups can be removed in parallel to oligonucleotide cleavage from the solid support to generate the oligonucleotide product, or the cyanoethyl group can be removed under milder condition before chain cleavage, if required.

[0009] Many aspects of the environmental impact of the current phosphoramidite methodology and potential advances have been reviewed (Andrews et al. (2021). J. Org. Chem. 86, 49–61). Even at an aspirational high oligonucleotide loading of 20% mass final oligonucleotide to mass solid support, at least five-fold mass of support is required over the final product mass.

[0010] In addition to the high cost of mass support required to immobilize the oligonucleotide, the use of organic solvents and 5'-O-blocking groups entail additional waste and process inefficiencies. Organic solvents such as acetonitrile are required for solubilization of the phosphoramidite coupling partners, or dichloromethane or toluene for deprotection steps. These solvents need to be anhydrous to reduce undesired hydrolysis of the phosphoramidite partners and can come from non-sustainable sources, adding cost, sustainability questions, and potential supply issues to the process.

[0011] The phosphoramidite coupling partners themselves carry a required blocking group at the 5'O-position, the nucleobase nitrogen atom (in A, C and G), and the nascent phosphate. The most common 5'O-blocking group, dimethoxytrityl, has a molecular mass of ~ 303 Da that approaches that of the heaviest native ribonucleotide fragment Gp with a mass of ~ 345 Da. This protecting group requires energy, resources, and effort to produce and append, and then requires disposal when separated from the desired materials.

[0012] In conclusion, a paradigm shift in oligonucleotide synthesis is necessary to enable siRNA therapeutics by lowering environmental impact, improving economic efficiency, and increasing scalability. New methods of oligonucleotide synthesis are, therefore, of great interest to the pharmaceutical industry.

Template-Independent Enzymatic Synthesis

[0013] Enzymatic synthesis may facilitate production of high volumes of complex or long polynucleotides (>200 base pairs) while minimizing toxic waste. A variety of prokaryotic and eukaryotic DNA and RNA polymerases are known to naturally synthesize polynucleotides of thousands of base pairs or more. Most of these polymerases function during DNA replication associated with cell division or transcription of RNA from DNA associated with gene or protein expression. Both of these processes involve template-dependent polynucleotide synthesis, wherein the polymerase uses an existing template polynucleotide strand to synthesize a complementary polynucleotide strand.

[0014] The potential of template-independent enzymatic polynucleotide synthesis to produce defined sequences has long been recognized. One early report suggested using NTPs with blocked 3' groups to allow stepwise addition of specific nucleotide residues (Bollum. (1962). JBC, 237, 1945–1949).

[0015] However, few polymerases are known to catalyze template-independent polynucleotide synthesis. These include polymerase lambda, polymerase mu, and terminal deoxynucleotidyl transferase (TdT), all members of the X family of DNA polymerases, many of which participate in DNA repair processes (Domínguez et al. (2000). EMBO, 19(7), 1731–1742.) Of these, TdT is known to generate diversity in antigen receptors by indiscriminately adding nucleosides to the 3' end of a single-stranded polynucleotide in a template-independent process (Bentolila et al. (1995). EMBO, 14(17), 4221–4229.)

[0016] Others have published a method of polynucleotide synthesis using a nucleoside 5'-triphosphate with a 3'-OH position protected with a removable blocking moiety and, specifically, a template-independent polynucleotide polymerase, including a terminal deoxynucleotidyl transferase (U.S. Pat. 5,763,594). The blocking group, also known to those skilled in the art as an inhibitor or reversible terminating group, may include a variety of groups that prevent the TdT from adding additional NTPs to the nascent polynucleotide chain. This may include charged molecules, large molecules and moieties, or other blocking groups known to those skilled in the art. Appropriate removable blocking groups may include carbonitriles, phosphates, carbonates, carbamates, esters, ethers, borates, nitrates, sugars, phosphoramidates, phenylsulfenates, and sulfates. Other 3' blocking groups are also known in the art, including 3'-O-amines and methylamines (U.S. Pat. 7,544,794) and 3'-O-azides (U.S. Pat. 10,407,721).

[0017] Although initially promising, use of 3'-blocked NTPs in template-independent synthesis catalyzed by TdT has proven difficult in practice, as TdT struggles to accept 3'-O-blocked NTPs as substrates. Further, wild-type TdTs have low tolerance for oligo acceptor substrates containing one or more modified nucleotides (e.g. 2' modifications).

[0018] Additionally, synthesis of RNA strands present unique challenges due to the additional, reactive 2'-OH on the ribose. While protection of the 2' position facilitates RNA synthesis, this approach reduces efficiency because of steric hindrance by the 2' protecting groups and requires maintenance and removal of the protecting group (CB Reese. (2005). Org Biomol Chem 3, 3851–3868.)

[0019] Recently several reports have described template-independent synthesis methods that use modified NTPs with blocking groups attached to the purine or pyrimidine base, leaving the 3'-OH unmodified and available for additional rounds of synthesis. These base blocking groups may include a cleavable linker that allows removal of the blocking group after each NTP addition step. The cleavable linker may also be attached to a detectable label (U.S. Pat. 7,057,026, among others). A variety of cleavable linkers are known to those skilled in the art. These include linkers attached via reducible disulfide bonds, photocleavable, electrophilic, or nucleophilic, pH sensitive, temperature sensitive, and linkers cleaved by enzymes. One drawback to using cleavable linkers is that, typically, some atoms of the linker moiety remain attached to the NTP following cleavage, leaving a "scar" that may interfere with synthesis of a complementary strand after initial template-independent synthesis of the primary polynucleotide strand.

[0020] Recently, modified NTPs with bases attached to blocking groups with cleavable linkers that are "scarless" and leave the nascent DNA ready for the next round of synthesis have been developed. In one example, the blocking group and cleavable linker are attached to the base via a disulfide bond. Upon addition of a reducing agent, the blocking group is removed, and the remaining atoms of the linker self-cyclize to leave the nascent DNA free of any linker atoms (U.S. Pat. 8,808,989, U.S. Pat. 9,695,470, U.S. Pat. 10,041,110). Methods of using NTPs attached to cleavable blocking groups to synthesize polynucleotides are known, including using a microfluidic device or ink jet printing technology (U.S. Pat. 9,279,149). An exonuclease may also be used in a method to synthesize polynucleotides to shorten or completely degrade polynucleotide strands that have not successfully added an NTP after the polynucleotide extension step and prior to removing the blocking group (U.S. Pat. 9,771,613).

[0021] However, NTP bases with bulky blocking groups attached via cleavable linkers are not optimal for efficient synthesis of complex or long oligonucleotides. The large labels may negatively impact enzyme kinetics, and linker scars may lead to an unacceptable rate of misincorporation when synthesizing the oligonucleotide strand. Additionally, larger linkers and necessary deblocking steps may increase the cost, time, and inefficiency of the process as a whole, rendering these methods economically infeasible.

[0022] Recently, several groups have explored modifying the structure or amino acid sequence of TdT or other polymerases to allow template-independent synthesis using 3'-O-blocked groups.

Efcavitch et al. describes incorporation of 3' modified dNTPs by TdT in template-independent synthesis using a murine or bacterial TdT with substituted amino acid residues (U.S. Pat. 10,059,929). Other reports describe engineered bovine and gar (*Lepisosteus oculatus*) TdTs that displayed improved activity over wild-type TdT (U.S. Pat. 10,745,727, PCT/GB2020/050247). Similarly, a variety of mutations have been described to improve the activity of Pol X family enzymes (WO 2017216472 A2). Finally, an N-terminal truncation of the BRCT domain (or alternatively mutation of the BRCT domain) of TdT has also been described as enhancing activity in the addition of reversibly blocked NTPs to the 3'-OH of a nucleic acid (US20210164008A1).

[0023] Feasible methods of template-independent enzymatic synthesis of complex or long polynucleotides have recognized commercial value. Recently, research efforts devoted to resolving challenges in this field have yielded improved engineered enzymes with activity in the template independent synthesis of polynucleotides, including DNA (PCT/US2022/078071) and RNA (PCT/US2023/076667).

[0024] In particular, TdT enzymes may enable the template independent synthesis of nucleic acid therapeutics, including therapeutic RNAs. Therapeutic RNAs, including siRNAs, ASOs, aptamers, and guide RNA for CRISPR-Cas systems, often include modifications to enhance stability, reduce off-target activity, reduce cellular toxicity, or target the therapeutic to specific cells or tissues. These include modifications to the 2' or 3' position of the sugar, modifications to the nucleobase, and modifications to the phosphate backbone.

[0025] Additionally, therapeutic RNAs may include addition of one or more conjugate moieties with useful properties. As an example, conjugate moieties have been used in existing therapeutic RNAs to direct the therapeutic RNA to a specific *in vivo* cellular target. These conjugate moieties may be added in a step wise manner during synthesis or may be added chemically in a post-synthesis step. In some cases, an NTP-3'-O-RBG or natural or modified NTP substrate comprising a reactive group may be incorporated into the RNA (or other polynucleotide). The reactive group may be used post-synthesis to add a conjugate moiety or other functional group. Similarly, a linker may be used on an NTP-3'-O-RBG or natural or modified NTP substrate as an attachment point for a conjugate moiety or reactive group.

[0026] Recent advances in template independent synthesis of long or complex polynucleotides, including therapeutic RNAs, highlight the potential to move beyond traditional phosphoramidite synthesis.

[0027] However, wild-type and engineered template-independent polymerases are not known to have activity in incorporating conjugate moieties, reactive groups, or linkers into an oligonucleotide strand.

[0028] In particular, enzymes capable of incorporating an NTP with a conjugate moiety, reactive group, or linker would be useful to target and/or further functionalize polynucleotides, including siRNAs.

SUMMARY

[0029] The present disclosure provides engineered terminal nucleotidyl transferase (TnT) polypeptides useful in template-independent polynucleotide synthesis, as well as compositions and methods of utilizing these engineered polypeptides. The TnTs of the present disclosure are variants of a predicted splice variant of the wild-type gene from *Monodelphis domestica* (SEQ ID NO: 10).

[0030] As used herein, the term terminal nucleotidyl transferase (TnT) is used to distinguish an engineered enzyme with activity on a variety of nucleoside triphosphates, including ribonucleoside triphosphates with 3' modifications or with 2' modifications or with both 2' and 3' modifications, from a wild-type TdT enzyme having wild-type TdT activity.

[0031] These engineered TnTs are also capable of adding nucleoside triphosphates with a 3' removable blocking group (NTP-3'-O-RBG) and other natural or modified NTPs to the 3'-OH end of a growing oligonucleotide or polynucleotide chain in a template-independent manner. After removal of the blocking group, additional rounds of NTP addition can be used to synthesize a polynucleotide with a defined sequence of bases without using a complementary template strand as a guide for NTP incorporation (template-independent synthesis).

[0032] In particular, the engineered TnTs of the present disclosure are capable of extending an oligo acceptor substrate or polynucleotide by addition of an NTP-3'-O-RBG or natural or modified NTP substrate comprising a conjugate moiety, reactive group, or linker. Similarly, the engineered TnTs described herein are capable of adding an NTP-3'-O-RBG or natural or modified NTP substrate onto an oligo acceptor substrate or polynucleotide comprising at least one conjugate moiety, reactive group, or linker.

[0033] The nucleotide comprising the conjugate moiety, reactive group, or linker may be at the 3' end, 5' end, or at an internal position of the RNA or other polynucleotide. The RNA or other polynucleotide may comprise one or more conjugate moieties, reactive groups, or linkers at one or more positions.

[0034] In some embodiments, the conjugate moiety comprises a saccharide, lectin, lipid, sterol, peptide, protein, antibody, aptamer, peptide nucleic acid, vitamin, toxin, hormone, or another molecule.

[0035] In some embodiments, the reactive group comprises an azido (N_3), amino (NH_3), bicyclo[6.1.0]nonyne (BCN), cyano (CN), dibenzocyclooctynyl, alkynyl, tetrazinyl, vinyl, or another

group. The reactive group may be useful to introduce further modifications to the RNA or other polynucleotide.

[0036] In some embodiments, the conjugate moiety or reactive group is attached to the nucleoside via a linker. In some embodiments, the linker is attached to the nucleobase or the sugar moiety of the nucleoside. In some embodiments, the linker comprises a cleavable linker.

[0037] In some embodiments, the present disclosure provides an engineered TnT polypeptide comprising an amino acid sequence having at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence identity to a reference sequence of SEQ ID NOs: 4, 6, 8, 10, and/or 12 and comprising at least one substitution or one substitution set at one or more positions, wherein the positions are numbered with reference to SEQ ID NOs: 4, 6, 8, 10, and/or 12 and wherein the engineered TnT polypeptide has increased specific activity on NTP-3'-O-RBG or natural or modified NTP substrates comprising a conjugate moiety, reactive group, or linker; increased incorporation efficiency in extension of oligo acceptor substrates comprising a conjugate moiety, reactive group, or linker; and/or increased activity on various oligo acceptor substrates comprising a conjugate moiety, reactive group, or linker as compared to a wild-type TnT (or TdT) or other TnTs or template-independent polymerases known to those of skill in the art. These engineered TnT polypeptides with one or more amino acid substitutions or substitution sets are described, below, in the detailed description of the invention.

[0038] In some additional embodiments, the engineered polypeptide comprises an amino acid sequence with at least 60% sequence identity to any even-numbered sequence set forth in SEQ ID NOs: 4, 6, 8, 10, and/or 12.

[0039] In some embodiments, the present disclosure provides a method of adding NTP-3'-O-RBGs or natural or modified NTP substrates comprising one or more conjugate moieties, reactive groups, or linkers onto a polynucleotide or oligo acceptor substrate. In some embodiments, the method further comprises an engineered TnT polypeptide comprising an amino acid sequence having at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence identity to a reference sequence of SEQ ID NOs: 4, 6, 8, 10, and/or 12 and comprising at least one substitution or one substitution set at one or more positions, wherein the positions are numbered with reference to SEQ ID NOs: 4, 6, 8, 10, and/or 12.

[0040] In some embodiments, the present disclosure provides a method of adding NTP-3'-O-RBGs or natural or modified NTP substrates onto a polynucleotide or oligo acceptor substrate containing one or more conjugate moieties, reactive groups, or linkers. In some embodiments, the method further comprises an engineered TnT polypeptide comprising an amino acid sequence having at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence identity to a reference sequence of SEQ ID NOs: 4, 6, 8, 10, and/or 12 and comprising at least one substitution or one

substitution set at one or more positions, wherein the positions are numbered with reference to SEQ ID NOs: 4, 6, 8, 10, and/or 12.

[0041] The present disclosure also provides an engineered polynucleotide encoding at least one engineered polypeptide described in the above paragraphs. In some embodiments, the engineered polynucleotide comprises the odd-numbered sequences set forth in SEQ ID NOs: 3, 5, 7, or 11.

[0042] The present disclosure further provides vectors comprising at least one engineered polynucleotide described above. In some embodiments, the vectors further comprise at least one control sequence.

[0043] The present disclosure also provides host cells comprising the vectors provided herein. In some embodiments, the host cell produces at least one engineered polypeptide provided herein.

[0044] The present disclosure further provides methods of producing an engineered TnT polypeptide, comprising the steps of culturing the host cell provided herein under conditions such that the engineered polynucleotide is expressed, and the engineered polypeptide is produced. In some embodiments, the methods further comprise the step of recovering the engineered polypeptide.

DESCRIPTION OF THE INVENTION

[0045] Unless defined otherwise, all technical and scientific terms used herein generally have the same meaning as commonly understood by one of ordinary skill in the art to which this invention pertains. Generally, the nomenclature used herein, and the laboratory procedures of cell culture, molecular genetics, microbiology, organic chemistry, analytical chemistry, and nucleic acid chemistry described below are those well-known and commonly employed in the art. Such techniques are well-known and described in numerous texts and reference works well known to those of skill in the art. Standard techniques, or modifications thereof, are used for chemical syntheses and chemical analyses. All patents, patent applications, articles and publications mentioned herein, both *supra* and *infra*, are hereby expressly incorporated herein by reference.

[0046] Although any suitable methods and materials similar or equivalent to those described herein find use in the practice of the present disclosure, some methods and materials are described herein. It is to be understood that this invention is not limited to the particular methodology, protocols, and reagents described, as these may vary, depending upon the context they are used by those of skill in the art. Accordingly, the terms defined immediately below are more fully described by reference to the invention as a whole.

[0047] It is to be understood that both the foregoing general description and the following detailed description are exemplary and explanatory only and are not restrictive of the present disclosure. The section headings used herein are for organizational purposes only and not to be construed as limiting the subject matter described. Numeric ranges are inclusive of the numbers defining the range. Thus,

every numerical range disclosed herein is intended to encompass every narrower numerical range that falls within such broader numerical range, as if such narrower numerical ranges were all expressly written herein. It is also intended that every maximum (or minimum) numerical limitation disclosed herein includes every lower (or higher) numerical limitation, as if such lower (or higher) numerical limitations were expressly written herein.

[0048] As used in this specification and the appended claims, the singular forms "a", "an" and "the" include plural referents unless the context clearly indicates otherwise. Thus, for example, reference to "a polypeptide" includes more than one polypeptide. Similarly, "comprise," "comprises," "comprising" "include," "includes," and "including" are interchangeable and not intended to be limiting.

[0049] It is to be understood that where descriptions of various embodiments use the term "comprising," those skilled in the art would understand that in some specific instances, an embodiment can be alternatively described using language "consisting essentially of" or "consisting of." It is to be further understood that where descriptions of various embodiments use the term "optional" or "optionally" the subsequently described event or circumstance may or may not occur, and that the description includes instances where the event or circumstance occurs and instances in which it does not. It is to be understood that both the foregoing general description, and the following detailed description are exemplary and explanatory only and are not restrictive of this disclosure. The section headings used herein are for organizational purposes only and not to be construed as limiting the subject matter described.

Abbreviations

[0050] The abbreviations used for the genetically encoded amino acids are conventional and are as follows:

Amino Acid	**Three-Letter**	**One-Letter Abbreviation**
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartate	Asp	D
Cysteine	Cys	C
Glutamate	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K

Amino Acid	Three-Letter	One-Letter Abbreviation
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

[0051] When the three-letter abbreviations are used, unless specifically preceded by an "L" or a "D" or clear from the context in which the abbreviation is used, the amino acid may be in either the L- or D-configuration about α-carbon (C_α). For example, whereas "Ala" designates alanine without specifying the configuration about the α-carbon, "D-Ala" and "L-Ala" designate D-alanine and L-alanine, respectively.

[0052] When the one-letter abbreviations are used, upper case letters designate amino acids in the L-configuration about the α-carbon and lower-case letters designate amino acids in the D-configuration about the α-carbon. For example, "A" designates L-alanine and "a" designates D-alanine. When polypeptide sequences are presented as a string of one-letter or three-letter abbreviations (or mixtures thereof), the sequences are presented in the amino (N) to carboxy (C) direction in accordance with common convention.

[0053] The abbreviations used for the genetically encoding nucleosides are conventional and are as follows: adenosine (A); guanosine (G); cytidine (C); thymidine (T); and uridine (U). These abbreviations are also used interchangeably for nucleosides and nucleotides (nucleosides with one or more phosphate groups). Unless specifically delineated, the abbreviated nucleosides or nucleotides may be either ribonucleosides (or ribonucleotides) or 2'-deoxyribonucleosides (or 2'-deoxyribonucleotides). The nucleosides or nucleotides may also be modified at the 3' position. The nucleosides or nucleotides may be specified as being either ribonucleosides (or ribonucleotides) or 2'-deoxyribonucleosides (or 2'-deoxyribonucleotides) on an individual basis or on an aggregate basis. When nucleic acid sequences are presented as a string of one-letter abbreviations, the sequences are presented in the 5' to 3' direction in accordance with common convention, and the phosphates are not indicated.

Definitions

[0054] In reference to the present disclosure, the technical and scientific terms used in the descriptions herein will have the meanings commonly understood by one of ordinary skill in the art,

unless specifically defined otherwise. Accordingly, the following terms are intended to have the following meanings.

[0055] "EC" number refers to the Enzyme Nomenclature of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB). The IUBMB biochemical classification is a numerical classification system for enzymes based on the chemical reactions they catalyze.

[0056] "ATCC" refers to the American Type Culture Collection whose biorepository collection includes genes and strains.

[0057] "NCBI" refers to National Center for Biological Information and the sequence databases provided therein.

[0058] "Protein," "polypeptide," and "peptide" are used interchangeably herein to denote a polymer of at least two amino acids covalently linked by an amide bond, regardless of length or post-translational modification (e.g., glycosylation, phosphorylation, lipidation, myristilation, ubiquitination, etc.). Included within this definition are D- and L-amino acids, and mixtures of D- and L-amino acids, as well as polymers comprising D- and L-amino acids, and mixtures of D- and L-amino acids.

[0059] "Amino acids" are referred to herein by either their commonly known three-letter symbols or by the one-letter symbols recommended by IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single letter codes.

[0060] As used herein, "polynucleotide," "oligonucleotide," and "nucleic acid'' are used interchangeably herein and refer to two or more nucleosides or nucleotides that are covalently linked together. The polynucleotide may be wholly comprised of ribonucleotides (i.e., RNA), wholly comprised of 2' deoxyribonucleotides (i.e., DNA), wholly comprised of other synthetic nucleotides or comprised of mixtures of synthetic, ribo- and/or 2' deoxyribonucleotides. The polynucleotides may also include modified nucleotides with substitutions, including 2' substitutions (e.g., 2'-fluoro, 2'-O-methyl, 2'-O-methoxyethyl, locked or constrained ethyl modifications, and others known to those skilled in the art). Nucleosides will be linked together via standard phosphodiester linkages or via one or more non-standard linkages, including but not limited to phosphorothioate linkages. The polynucleotide may be single-stranded or double-stranded or may include both single-stranded regions and double-stranded regions. Moreover, while a polynucleotide will typically be composed of the naturally occurring encoding nucleobases (i.e., adenine, guanine, uracil, thymine, and cytosine), it may include one or more modified and/or synthetic nucleobases, such as, for example, inosine, xanthine, hypoxanthine, etc. In some embodiments, such modified or synthetic nucleobases are nucleobases encoding amino-acid sequences. Nucleobases that are modified or synthetic may comprise any known or hypothetical or future discovered modification or structure that would be

recognized by one of skill in the art as a modified or synthetic nucleobase. Similarly, the terms "polynucleotide," "oligonucleotide," and "nucleic acid'' are intended to comprise any modified or synthetic structure that is now known or discovered in the future that would be recognized by one of skill in the art as being or having the function of a "polynucleotide," "oligonucleotide," or "nucleic acid.'' An example of a modified or synthetic structure having the function of a "polynucleotide," "oligonucleotide," or "nucleic acid'' is PNA or peptide nucleic acid.

[0061] As used herein, "oligo acceptor substrate" and "acceptor substrate" and "growing oligo acceptor substrate strand" and "growing oligonucleotide chain" and "growing polynucleotide strand" are used interchangeably herein and refer to any oligo or nucleotide chain or similar moiety with an exposed 3'-OH or equivalent thereof that may be recognized by a wild-type TnT (or TdT) or polymerase or an engineered TnT or template-independent polymerase of the current disclosure as a substrate for nucleoside addition or synthesis. In some embodiments, the acceptor substrate may be single stranded. In yet other embodiments, the acceptor substrate may be double stranded or partially doubled stranded. In some embodiments, the acceptor substrate may comprise a nucleotide chain consisting of 1-10 nucleotides, 5-20 nucleotides, 15-50 nucleotides, 30-100 nucleotides, or greater than 100 nucleotides. In some embodiments, the acceptor substrate may comprise a chemical moiety that is not a nucleotide chain but contains a free -OH capable of being recognized as a substrate by a wild-type or engineered TnT, referred to herein as a "3'-OH equivalent". Exemplary oligo acceptor substrates are provided in the Examples.

[0062] As used herein, "nucleoside triphosphate-3'-O-removable blocking group" and "nucleotide triphosphate-3'-O-removable blocking group" and "reversible terminator" and "NTP-3'-O-RBG" are used interchangeably herein and refer to a ribonucleoside triphosphate or a deoxyribonucleoside triphosphate or a synthetic or nucleoside triphosphate composed of an alternate or modified sugar with a removable blocking group attached at the 3' position of the sugar moiety. An NTP-3'-O-RBG may also include other modifications as described herein, including but not limited to modifications at the 2' position, modifications to the nucleobase, and modifications to the phosphates. A nucleotide may also have a 3'-O-RBG, as is expected after reaction of an NTP-3'-O-RBG with an engineered TnT of the present disclosure and an oligo acceptor substrate.

[0063] As used herein, "oligo acceptor product" and "growing oligonucleotide chain" and "oligo acceptor extension product" are used interchangeably herein and refer to the product of a NTP-3'-O-RBG or other natural or modified NTP substrate and an oligo acceptor substrate, wherein a TnT or related polymerase has catalyzed the extension or addition of a nucleotide-3'-O-RBG or other natural or modified nucleotide substrate to an oligo acceptor substrate via reaction with one or more NTP-3'-O-RBGs or other natural or modified NTP substrates.

[0064] As used herein, "removable blocking group" and "blocking group" and "terminator group" and "reversible terminating group" and "inhibitor group" and related variations of these terms are used interchangeably herein and refer to a chemical group that would hinder addition of a second NTP-3'-O-RBG or other natural or modified NTP substrate to the 3' end of the growing oligo acceptor substrate strand prior to removal of the removable blocking from the first round of addition. In some embodiments, the NTP-3'-O-RBG or other natural or modified NTP substrate may comprise a removable blocking group selected from the group consisting of NTP-3'-O-NH_2, or NTP-3'-O-PO_3. In some embodiments, the NTP- 3'-O-RBG or other natural or modified NTP substrate may have a natural purine or pyrimidine base, such as adenine, guanine, cytosine, thymine, or uridine. In some embodiments, NTP- 3'-O-RBG or other natural or modified NTP substrates may have an unnatural base analog such as inosine, xanthine, hypoxanthine, or another base analog, as is known in the art. In some embodiments the blocking group may comprise or may additionally comprise a modification at the 2' position.

[0065] As used herein, "template-independent synthesis" refers to synthesis of an oligonucleotide or a polynucleotide without the use of template strand as a guide for synthesis of a complementary oligo or polynucleotide strand. Thus, template-independent synthesis refers to an iterative process, whereby, successive nucleotides are added to a growing oligo or nucleotide chain or acceptor substrate. Template-independent synthesis may be in a sequence defined manner or may be random, as is the case with the wild-type TnT (or TdT) in creating antigen receptor diversity. Processes for template-independent synthesis are further described herein.

[0066] "Coding sequence" refers to that portion of a nucleic acid (e.g., a gene) that encodes an amino acid sequence of a protein.

[0067] "Naturally-occurring" or "wild-type" refers to the form found in nature. For example, a naturally occurring or wild-type polypeptide or polynucleotide sequence is a sequence present in an organism that can be isolated from a source in nature and which has not been intentionally modified by human manipulation.

[0068] As used herein, "recombinant," "engineered," and "non-naturally occurring" when used with reference to a cell, nucleic acid, or polypeptide, refer to a material, or a material corresponding to the natural or native form of the material, that has been modified in a manner that would not otherwise exist in nature. In some embodiments, the cell, nucleic acid, or polypeptide is identical to a naturally occurring cell, nucleic acid, or polypeptide, but is produced or derived from synthetic materials and/or by manipulation using recombinant techniques. Non-limiting examples include, among others, recombinant cells expressing genes that are not found within the native (non-recombinant) form of the cell or expressed native genes that are otherwise expressed at a different level.

[0069] "Percentage of sequence identity" and "percentage homology" are used interchangeably herein to refer to comparisons among polynucleotides or polypeptides and are determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide or polypeptide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence for optimal alignment of the two sequences. The percentage may be calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity. Alternatively, the percentage may be calculated by determining the number of positions at which either the identical nucleic acid base or amino acid residue occurs in both sequences or a nucleic acid base or amino acid residue is aligned with a gap to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity. Those of skill in the art appreciate that there are many established algorithms available to align two sequences. Optimal alignment of sequences for comparison can be conducted, *e.g.*, by the local homology algorithm of Smith and Waterman (Smith and Waterman, Adv. Appl. Math., 2:482 [1981]), by the homology alignment algorithm of Needleman and Wunsch (Needleman and Wunsch, J. Mol. Biol., 48:443 [1970]), by the search for similarity method of Pearson and Lipman (Pearson and Lipman, Proc. Natl. Acad. Sci. USA 85:2444 [1988]), by computerized implementations of these algorithms (e.g., GAP, BESTFIT, FASTA, and TFASTA in the GCG Wisconsin Software Package), or by visual inspection, as known in the art. Examples of algorithms that are suitable for determining percent sequence identity and sequence similarity include, but are not limited to the BLAST and BLAST 2.0 algorithms, which are described by Altschul et al. (See, Altschul et al., J. Mol. Biol., 215: 403-410 [1990]; and Altschul et al., Nucl. Acids Res., 3389-3402 [1977], respectively). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information website. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as, the neighborhood word score threshold (Altschul et al., *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always >0) and N (penalty score for mismatching residues; always <0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the

cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSum62 scoring matrix (See, Henikoff and Henikoff, Proc. Natl. Acad. Sci. USA 89:10915 [1989]). Exemplary determination of sequence alignment and % sequence identity can employ the BESTFIT or GAP programs in the GCG Wisconsin Software package (Accelrys, Madison WI), using default parameters provided.

[0070] "Reference sequence" refers to a defined sequence used as a basis for a sequence comparison. A reference sequence may be a subset of a larger sequence, for example, a segment of a full-length gene or polypeptide sequence. Generally, a reference sequence is at least 20 nucleotide or amino acid residues in length, at least 25 residues in length, at least 50 residues in length, or the full length of the nucleic acid or polypeptide. Since two polynucleotides or polypeptides may each (1) comprise a sequence (*i.e.*, a portion of the complete sequence) that is similar between the two sequences, and (2) may further comprise a sequence that is divergent between the two sequences, sequence comparisons between two (or more) polynucleotides or polypeptide are typically performed by comparing sequences of the two polynucleotides or polypeptides over a "comparison window" to identify and compare local regions of sequence similarity. In some embodiments, a "reference sequence" can be based on a primary amino acid sequence, where the reference sequence is a sequence that can have one or more changes in the primary sequence. For instance, a "reference sequence based on SEQ ID NO:4 having at the residue corresponding to X14 a valine" or X14V refers to a reference sequence in which the corresponding residue at X14 in SEQ ID NO:4, which is a tyrosine, has been changed to valine.

[0071] "Comparison window" refers to a conceptual segment of at least about 20 contiguous nucleotide positions or amino acids residues wherein a sequence may be compared to a reference sequence of at least 20 contiguous nucleotides or amino acids and wherein the portion of the sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) of 20 percent or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The comparison window can be longer than 20 contiguous residues, and includes, optionally 30, 40, 50, 100, or longer windows.

[0072] As used herein, "substantial identity" refers to a polynucleotide or polypeptide sequence that has at least 80 percent sequence identity, at least 85 percent identity, at least between 89 to 95 percent sequence identity, or more usually, at least 99 percent sequence identity as compared to a reference sequence over a comparison window of at least 20 residue positions, frequently over a window of at least 30-50 residues, wherein the percentage of sequence identity is calculated by comparing the

reference sequence to a sequence that includes deletions or additions which total 20 percent or less of the reference sequence over the window of comparison. In some specific embodiments applied to polypeptides, the term "substantial identity" means that two polypeptide sequences, when optimally aligned, such as by the programs GAP or BESTFIT using default gap weights, share at least 80 percent sequence identity, preferably at least 89 percent sequence identity, at least 95 percent sequence identity or more (e.g., 99 percent sequence identity). In some embodiments, residue positions that are not identical in sequences being compared differ by conservative amino acid substitutions.

[0073] "Corresponding to," "reference to," and "relative to" when used in the context of the numbering of a given amino acid or polynucleotide sequence refer to the numbering of the residues of a specified reference sequence when the given amino acid or polynucleotide sequence is compared to the reference sequence. In other words, the residue number or residue position of a given polymer is designated with respect to the reference sequence rather than by the actual numerical position of the residue within the given amino acid or polynucleotide sequence. For example, a given amino acid sequence, such as that of an engineered TnT, can be aligned to a reference sequence by introducing gaps to optimize residue matches between the two sequences. In these cases, although the gaps are present, the numbering of the residue in the given amino acid or polynucleotide sequence is made with respect to the reference sequence to which it has been aligned.

[0074] "Amino acid difference" or "residue difference" refers to a change in the amino acid residue at a position of a polypeptide sequence relative to the amino acid residue at a corresponding position in a reference sequence. The positions of amino acid differences generally are referred to herein as "Xn," where n refers to the corresponding position in the reference sequence upon which the residue difference is based. For example, a "residue difference at position X25 as compared to SEQ ID NO: 2" refers to a change of the amino acid residue at the polypeptide position corresponding to position 25 of SEQ ID NO:2. Thus, if the reference polypeptide of SEQ ID NO: 2 has a valine at position 25, then a "residue difference at position X25 as compared to SEQ ID NO:2" an amino acid substitution of any residue other than valine at the position of the polypeptide corresponding to position 25 of SEQ ID NO: 2. In most instances herein, the specific amino acid residue difference at a position is indicated as "XnY" where "Xn" specified the corresponding position as described above, and "Y" is the single letter identifier of the amino acid found in the engineered polypeptide (i.e., the different residue than in the reference polypeptide). In some embodiments, more than one amino acid can appear in a specified residue position (i.e., the alternative amino acids can be listed in the form XnY/Z, where Y and Z represent alternate amino acid residues). In some instances, the present disclosure also provides specific amino acid differences denoted by the conventional notation "AnB", where A is the single letter identifier of the residue in the reference sequence, "n" is the number of the residue position in the reference sequence, and B is the single letter identifier of the residue

substitution in the sequence of the engineered polypeptide. Furthermore, in some instances, a polypeptide of the present disclosure can include one or more amino acid residue differences relative to a reference sequence, which is indicated by a list of the specified positions where changes are made relative to the reference sequence. In some additional embodiments, the present disclosure provides engineered polypeptide sequences comprising both conservative and non-conservative amino acid substitutions.

[0075] As used herein, "conservative amino acid substitution" refers to a substitution of a residue with a different residue having a similar side chain, and thus typically involves substitution of the amino acid in the polypeptide with amino acids within the same or similar defined class of amino acids. By way of example and not limitation, an amino acid with an aliphatic side chain is substituted with another aliphatic amino acid (e.g., alanine, valine, leucine, and isoleucine); an amino acid with an hydroxyl side chain is substituted with another amino acid with a hydroxyl side chain (e.g., serine and threonine); an amino acid having an aromatic side chain is substituted with another amino acid having an aromatic side chain (e.g., phenylalanine, tyrosine, tryptophan, and histidine); an amino acid with a basic side chain is substituted with another amino acid with a basic side chain (e.g., lysine and arginine); an amino acid with an acidic side chain is substituted with another amino acid with an acidic side chain (e.g., aspartic acid or glutamic acid); and/or a hydrophobic or hydrophilic amino acid is replaced with another hydrophobic or hydrophilic amino acid, respectively. Exemplary conservative substitutions are provided in Table 1 below.

Table 1. Conservative Amino Acid Substitution Examples

Residue	Possible Conservative Substitutions
A, L, V, I	Other aliphatic (A, L, V, I) Other non-polar (A, L, V, I, G, M)
G, M	Other non-polar (A, L, V, I, G, M)
D, E	Other acidic (D, E)
K, R	Other basic (K, R)
N, Q, S, T	Other polar
H, Y, W, F	Other aromatic (H, Y, W, F)
C, P	None

[0076] "Non-conservative substitution" refers to substitution of an amino acid in the polypeptide with an amino acid with significantly differing side chain properties. Non-conservative substitutions may use amino acids between, rather than within, the defined groups and affects (a) the structure of the peptide backbone in the area of the substitution (*e.g.*, proline for glycine), (b) the charge or hydrophobicity, or (c) the bulk of the side chain. By way of example and not limitation, an exemplary non-conservative substitution can be an acidic amino acid substituted with a basic or aliphatic amino acid; an aromatic amino acid substituted with a small amino acid; and a hydrophilic amino acid substituted with a hydrophobic amino acid.

[0077] "Deletion" refers to modification to the polypeptide by removal of one or more amino acids from the reference polypeptide. Deletions can comprise removal of 1 or more amino acids, 2 or more amino acids, 5 or more amino acids, 10 or more amino acids, 15 or more amino acids, or 20 or more amino acids, up to 10% of the total number of amino acids, or up to 20% of the total number of amino acids making up the reference enzyme while retaining enzymatic activity and/or retaining the improved properties of an engineered TnT enzyme. Deletions can be directed to the internal portions and/or terminal portions of the polypeptide. In various embodiments, the deletion can comprise a continuous segment or can be discontinuous.

[0078] "Insertion" refers to modification to the polypeptide by addition of one or more amino acids from the reference polypeptide. In some embodiments, the improved engineered TnT enzymes comprise insertions of one or more amino acids to the naturally occurring polypeptide as well as insertions of one or more amino acids to other improved TnT polypeptides. Insertions can be in the internal portions of the polypeptide, or to the carboxy or amino terminus. Insertions as used herein include fusion proteins as is known in the art. The insertion can be a contiguous segment of amino acids or separated by one or more of the amino acids in the naturally occurring polypeptide.

[0079] "Fragment" as used herein refers to a polypeptide that has an amino-terminal and/or carboxy-terminal deletion, but where the remaining amino acid sequence is identical to the corresponding positions in the sequence. Fragments can be at least 14 amino acids long, at least 20 amino acids long, at least 50 amino acids long or longer, and up to 70%, 80%, 90%, 95%, 98%, and 99% of the full-length TnT polypeptide, for example the polypeptide of SEQ ID NO: 4 or an TnT provided in the even-numbered sequences of SEQ ID NO: 4-12.

[0080] "Isolated polypeptide" refers to a polypeptide which is substantially separated from other contaminants that naturally accompany it, *e.g.*, protein, lipids, and polynucleotides. The term embraces polypeptides which have been removed or purified from their naturally-occurring environment or expression system (*e.g.*, host cell or *in vitro* synthesis). The engineered TnT enzymes may be present within a cell, present in the cellular medium, or prepared in various forms, such as lysates or isolated preparations. As such, in some embodiments, the engineered TnT enzyme can be an isolated polypeptide.

[0081] "Substantially pure polypeptide" refers to a composition in which the polypeptide species is the predominant species present (*i.e.*, on a molar or weight basis it is more abundant than any other individual macromolecular species in the composition) and is generally a substantially purified composition when the object species comprises at least about 50 percent of the macromolecular species present by mole or % weight. Generally, a substantially pure TnT composition will comprise about 60 % or more, about 70% or more, about 80% or more, about 90% or more, about 95% or more, and about 98% or more of all macromolecular species by mole or % weight present in the

composition. In some embodiments, the object species is purified to essential homogeneity (*i.e.*, contaminant species cannot be detected in the composition by conventional detection methods) wherein the composition consists essentially of a single macromolecular species. Solvent species, small molecules (<500 Daltons), and elemental ion species are not considered macromolecular species. In some embodiments, the isolated engineered TnT polypeptide is a substantially pure polypeptide composition.

[0082] As used herein, "improved enzyme property" refers to at least one improved property of an enzyme. In some embodiments, the present disclosure provides engineered TnT polypeptides that exhibit an improvement in any enzyme property as compared to a reference TnT polypeptide and/or a wild-type TnT (or TdT) polypeptide, and/or another engineered TnT polypeptide. For the engineered TnT polypeptides described herein, the comparison is generally made to the wild-type enzyme from which the TnT is derived, although in some embodiments, the reference enzyme can be another improved engineered TnT. Thus, the level of "improvement" can be determined and compared between various TnT polypeptides, including wild-type, as well as engineered TnTs. Improved properties include, but are not limited, to such properties as enzymatic activity (which can be expressed in terms of percent conversion of the substrate), thermostability, solvent stability, pH activity profile, cofactor requirements, refractoriness to inhibitors (*e.g.*, substrate or product inhibition), activity at elevated temperatures, increased soluble expression, decreased by-product formation, increased specific activity on NTP-3'-O-RBG substrates, increased incorporation efficiency in extension of oligo acceptor substrates, and/or increased activity on various oligo acceptor substrates (including enantioselectivity).

[0083] "Increased enzymatic activity" refers to an improved property of the TnT polypeptides, which can be represented by an increase in specific activity (*e.g.*, product produced/time/weight protein) or an increase in percent conversion of the substrate to the product (*e.g.*, percent conversion of starting amount of substrate to product in a specified time period using a specified amount of TnT) as compared to the reference TnT enzyme. Exemplary methods to determine enzyme activity are provided in the Examples. Any property relating to enzyme activity may be affected, including the classical enzyme properties of K_m, V_{max} or k_{cat}, changes of which can lead to increased enzymatic activity. Improvements in enzyme activity can be from about 1.2 times the enzymatic activity of the corresponding wild-type enzyme, to as much as 2 times, 5 times, 10 times, 20 times, 25 times, 50 times or more enzymatic activity than the naturally occurring or another engineered TnT from which the TnT polypeptides were derived. TnT activity can be measured by any one of standard assays, such as by monitoring changes in properties of substrates, cofactors, or products. In some embodiments, the amount of products generated can be measured by Liquid Chromatography-Mass Spectrometry (LC-MS), HPLC, or other methods, as known in the art. Comparisons of enzyme activities are made using a defined preparation of enzyme, a defined assay under a set condition, and

one or more defined substrates, as further described in detail herein. Generally, when lysates are compared, the numbers of cells and the amount of protein assayed are determined as well as use of identical expression systems and identical host cells to minimize variations in amount of enzyme produced by the host cells and present in the lysates.

[0084] "Conversion" refers to the enzymatic conversion of the substrate(s) to the corresponding product(s). "Percent conversion" refers to the percent of the substrate that is converted to the product within a period of time under specified conditions. Thus, the "enzymatic activity" or "activity" of a TnT polypeptide can be expressed as "percent conversion" of the substrate to the product.

[0085] "Thermostable" refers to a polypeptide that maintains similar activity (more than 60% to 80% for example) after exposure to elevated temperatures (*e.g.*, 40-80 °C) for a period of time (*e.g.*, 0.5-24 hrs) compared to the wild-type enzyme exposed to the same elevated temperature.

[0086] "Solvent stable" refers to a polypeptide that maintains similar activity (more than *e.g.*, 60% to 80%) after exposure to varying concentrations (*e.g.*, 5-99%) of solvent (ethanol, isopropyl alcohol, dimethylsulfoxide (DMSO), tetrahydrofuran, 2-methyltetrahydrofuran, acetone, toluene, butyl acetate, methyl tert-butyl ether, etc.) for a period of time (*e.g.*, 0.5-24 hrs) compared to the wild-type enzyme exposed to the same concentration of the same solvent.

[0087] "Thermo- and solvent stable" refers to a polypeptide that is both thermostable and solvent stable.

[0088] The term "stringent hybridization conditions" is used herein to refer to conditions under which nucleic acid hybrids are stable. As known to those of skill in the art, the stability of hybrids is reflected in the melting temperature (T_m) of the hybrids. In general, the stability of a hybrid is a function of ion strength, temperature, G/C content, and the presence of chaotropic agents. The T_m values for polynucleotides can be calculated using known methods for predicting melting temperatures (See e.g., Baldino et al., Meth. Enzymol., 168:761-777 [1989]; Bolton et al., Proc. Natl. Acad. Sci. USA 48:1390 [1962]; Breslauer et al., Proc. Natl. Acad. Sci. USA 83:8893-8897 [1983]; Freier et al., Proc. Natl. Acad. Sci. USA 83:9373-9377 [1986]; Kierzek et al., Biochem., 25:7840-7846 [1986]; Rychlik et al., 1990, Nucl. Acids Res., 18:6409-6412 [1990] (erratum, Nucl. Acids Res., 19:698 [1991]); Suggs et al., 1981, *in* Developmental Biology Using Purified Genes, Brown et al. [eds.], pp. 683-693, Academic Press, Cambridge, MA [1981]; and Wetmur, Crit. Rev. Biochem. Mol. Biol., 26:227-259 [1991]). In some embodiments, the polynucleotide encodes the polypeptide disclosed herein and hybridizes under defined conditions, such as moderately stringent or highly stringent conditions, to the complement of a sequence encoding an engineered TnT enzyme of the present disclosure.

[0089] "Hybridization stringency" relates to hybridization conditions, such as washing conditions, in the hybridization of nucleic acids. Generally, hybridization reactions are performed under conditions

of lower stringency, followed by washes of varying but higher stringency. The term "moderately stringent hybridization" refers to conditions that permit target-DNA to bind a complementary nucleic acid that has about 60% identity, preferably about 75% identity, about 85% identity to the target DNA, with greater than about 90% identity to target-polynucleotide. Exemplary moderately stringent conditions are conditions equivalent to hybridization in 50% formamide, 5× Denhart's solution, 5×SSPE, 0.2% SDS at 42 °C, followed by washing in 0.2×SSPE, 0.2% SDS, at 42 °C. "High stringency hybridization" refers generally to conditions that are about 10 °C or less from the thermal melting temperature T_m as determined under the solution condition for a defined polynucleotide sequence. In some embodiments, a high stringency condition refers to conditions that permit hybridization of only those nucleic acid sequences that form stable hybrids in 0.018M NaCl at 65 °C (*i.e.*, if a hybrid is not stable in 0.018M NaCl at 65 °C, it will not be stable under high stringency conditions, as contemplated herein). High stringency conditions can be provided, for example, by hybridization in conditions equivalent to 50% formamide, 5× Denhart's solution, 5×SSPE, 0.2% SDS at 42 °C, followed by washing in 0.1×SSPE, and 0.1% SDS at 65 °C. Another high stringency condition is hybridizing in conditions equivalent to hybridizing in 5X SSC containing 0.1% (w:v) SDS at 65 °C and washing in 0.1x SSC containing 0.1% SDS at 65 °C. Other high stringency hybridization conditions, as well as moderately stringent conditions, are described in the references cited above.

[0090] "Heterologous" polynucleotide refers to any polynucleotide that is introduced into a host cell by laboratory techniques and includes polynucleotides that are removed from a host cell, subjected to laboratory manipulation, and then reintroduced into a host cell.

[0091] "Codon optimized" refers to changes in the codons of the polynucleotide encoding a protein to those preferentially used in a particular organism such that the encoded protein is efficiently expressed in the organism of interest. Although the genetic code is degenerate in that most amino acids are represented by several codons, called "synonyms" or "synonymous" codons, it is well known that codon usage by particular organisms is nonrandom and biased towards particular codon triplets. This codon usage bias may be higher in reference to a given gene, genes of common function or ancestral origin, highly expressed proteins versus low copy number proteins, and the aggregate protein coding regions of an organism's genome. In some embodiments, the polynucleotides encoding the TnT enzymes may be codon optimized for optimal production from the host organism selected for expression.

[0092] As used herein, "preferred, optimal, high codon usage bias codons" refers interchangeably to codons that are used at higher frequency in the protein coding regions than other codons that code for the same amino acid. The preferred codons may be determined in relation to codon usage in a single gene, a set of genes of common function or origin, highly expressed genes, the codon frequency in the aggregate protein coding regions of the whole organism, codon frequency in the aggregate protein

coding regions of related organisms, or combinations thereof. Codons whose frequency increases with the level of gene expression are typically optimal codons for expression. A variety of methods are known for determining the codon frequency (e.g., codon usage, relative synonymous codon usage) and codon preference in specific organisms, including multivariate analysis, for example, using cluster analysis or correspondence analysis, and the effective number of codons used in a gene (See e.g., GCG CodonPreference, Genetics Computer Group Wisconsin Package; CodonW, Peden, University of Nottingham; McInerney, Bioinform., 14:372-73 [1998]; Stenico et al., Nucl. Acids Res., 222437-46 [1994]; Wright, Gene 87:23-29 [1990]). Codon usage tables are available for many different organisms (See e.g., Wada et al., Nucl. Acids Res., 20:2111-2118 [1992]; Nakamura et al., Nucl. Acids Res., 28:292 [2000]; Henaut and Danchin, *in* *Escherichia coli* and *Salmonella*, Neidhardt, et al. (eds.), ASM Press, Washington D.C., p. 2047-2066 [1996]). The data source for obtaining codon usage may rely on any available nucleotide sequence capable of coding for a protein. These data sets include nucleic acid sequences actually known to encode expressed proteins (e.g., complete protein coding sequences-CDS), expressed sequence tags (ESTS), or predicted coding regions of genomic sequences (See e.g., Mount, Bioinformatics: Sequence and Genome Analysis, Chapter 8, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. [2001]; Uberbacher, Meth. Enzymol., 266:259-281 [1996]; and Tiwari et al., Comput. Appl. Biosci., 13:263-270 [1997]).

[0093] "Control sequence" is defined herein to include all components, which are necessary or advantageous for the expression of a polynucleotide and/or polypeptide of the present disclosure. Each control sequence may be native or foreign to the nucleic acid sequence encoding the polypeptide. Such control sequences include, but are not limited to, a leader, polyadenylation sequence, propeptide sequence, promoter, signal peptide sequence, and transcription terminator. At a minimum, the control sequences include a promoter, and transcriptional and translational stop signals. The control sequences may be provided with linkers for the purpose of introducing specific restriction sites facilitating ligation of the control sequences with the coding region of the nucleic acid sequence encoding a polypeptide.

[0094] "Operably linked" is defined herein as a configuration in which a control sequence is appropriately placed (*i.e.*, in a functional relationship) at a position relative to a polynucleotide of interest such that the control sequence directs or regulates the expression of the polynucleotide and/or polypeptide of interest.

[0095] "Promoter sequence" refers to a nucleic acid sequence that is recognized by a host cell for expression of a polynucleotide of interest, such as a coding sequence. The promoter sequence contains transcriptional control sequences, which mediate the expression of a polynucleotide of interest. The promoter may be any nucleic acid sequence which shows transcriptional activity in the host cell of choice including mutant, truncated, and hybrid promoters, and may be obtained from genes encoding extracellular or intracellular polypeptides either homologous or heterologous to the host cell.

[0096] "Suitable reaction conditions" refer to those conditions in the biocatalytic reaction solution (*e.g.*, ranges of enzyme loading, substrate loading, cofactor loading, temperature, pH, buffers, co-solvents, etc.) under which a TnT polypeptide of the present disclosure is capable of converting one or more substrate compounds to a product compound (*e.g.*, addition of a nucleotide-3'-O-RBG or other natural or modified nucleotide substrate to an oligo acceptor substrate via reaction with NTP-3'-O-RBG or other natural or modified NTP substrate). Exemplary "suitable reaction conditions" are provided in the present disclosure and illustrated by the Examples.

[0097] "Composition" refers to a mixture or combination of one or more substances, wherein each substance or component of the composition retains its individual properties. As used herein, a biocatalytic composition refers to a combination of one or more substances useful for biocatalysis.

[0098] "Loading", such as in "compound loading" or "enzyme loading" or "cofactor loading" refers to the concentration or amount of a component in a reaction mixture at the start of the reaction.

[0099] "Substrate" in the context of a biocatalyst mediated process refers to the compound or molecule acted on by the biocatalyst. For example, a TnT biocatalyst used in the synthesis processes disclosed herein acts on an NTP-3'-O-RBG substrate or other natural or modified NTP substrate and an oligo acceptor substrate.

[0100] "Product" in the context of a biocatalyst mediated process refers to the compound or molecule resulting from the action of the biocatalyst. For example, an exemplary product for a TnT biocatalyst used in a process disclosed herein is an oligo acceptor extension product, as depicted in Schemes 1 and 2.

[0101] "Alkyl" refers to saturated hydrocarbon groups of from 1 to 18 carbon atoms inclusively, either straight chained or branched, more preferably from 1 to 8 carbon atoms inclusively, and most preferably 1 to 6 carbon atoms inclusively. An alkyl with a specified number of carbon atoms is denoted in parenthesis (e.g., $(C_1\text{-}C_6)$alkyl refers to an alkyl of 1 to 6 carbon atoms).

[0102] "Alkenyl" refers to hydrocarbon groups of from 2 to 12 carbon atoms inclusively, either straight or branched containing at least one double bond but optionally containing more than one double bond.

[0103] "Alkynyl" refers to hydrocarbon groups of from 2 to 12 carbon atoms inclusively, either straight or branched containing at least one triple bond but optionally containing more than one triple bond, and additionally optionally containing one or more double bonded moieties.

[0104] "Heteroalkyl, "heteroalkenyl," and heteroalkynyl," refer respectively, to alkyl, alkenyl and alkynyl as defined herein in which one or more of the carbon atoms are each independently replaced with the same or different heteroatoms or heteroatomic groups. Heteroatoms and/or heteroatomic groups which can replace the carbon atoms include, but are not limited to -O-, -S-, -S-O-, -NR^{Y}-, -PH-,

-S(O)-, -S(O)2-, -S(O) NR^{γ}-, -$S(O)_2NR^{\gamma}$, and the like, including combinations thereof, where each R^{γ} is independently selected from hydrogen, alkyl, cycloalkyl, heterocycloalkyl, aryl, and heteroaryl.

[0105] "Amino" refers to the group -NH_2. Substituted amino refers to the group –NHR^{η}, $NR^{\eta}R^{\eta}$, and $NR^{\eta}R^{\eta}R^{\eta}$, where each R^{η} is independently selected from substituted or unsubstituted alkyl, cycloalkyl, cycloheteroalkyl, alkoxy, aryl, heteroaryl, heteroarylalkyl, acyl, alkoxycarbonyl, sulfanyl, sulfinyl, sulfonyl, and the like. Typical amino groups include, but are limited to, dimethylamino, diethylamino, trimethylammonium, triethylammonium, methylysulfonylamino, furanyl-oxy-sulfamino, and the like.

[0106] "Aminoalkyl" refers to an alkyl group in which one or more of the hydrogen atoms are replaced with one or more amino groups, including substituted amino groups.

[0107] "Aminocarbonyl" refers to -$C(O)NH_2$. Substituted aminocarbonyl refers to –$C(O)NR^{\eta}R^{\eta}$, where the amino group $NR^{\eta}R^{\eta}$ is as defined herein.

[0108] "Oxy" refers to a divalent group -O-, which may have various substituents to form different oxy groups, including ethers and esters.

[0109] "Alkoxy" or "alkyloxy" are used interchangeably herein to refer to the group –OR^{ζ}, wherein R^{ζ} is an alkyl group, including optionally substituted alkyl groups.

[0110] "Carboxy" refers to -COOH.

[0111] "Carbonyl" refers to -C(O)-, which may have a variety of substituents to form different carbonyl groups including acids, acid halides, aldehydes, amides, esters, and ketones.

[0112] "Carboxyalkyl" refers to an alkyl in which one or more of the hydrogen atoms are replaced with one or more carboxy groups.

[0113] "Aminocarbonylalkyl" refers to an alkyl substituted with an aminocarbonyl group, as defined herein.

[0114] "Halogen" or "halo" refers to fluoro, chloro, bromo and iodo.

[0115] "Haloalkyl" refers to an alkyl group in which one or more of the hydrogen atoms are replaced with a halogen. Thus, the term "haloalkyl" is meant to include monohaloalkyls, dihaloalkyls, trihaloalkyls, etc. up to perhaloalkyls. For example, the expression "(C_1 - C_2) haloalkyl" includes 1-fluoromethyl, difluoromethyl, trifluoromethyl, 1-fluoroethyl, 1,1-difluoroethyl, 1,2-difluoroethyl, 1,1,1 trifluoroethyl, perfluoroethyl, etc.

[0116] "Hydroxy" refers to -OH.

[0117] "Hydroxyalkyl" refers to an alkyl group in which in which one or more of the hydrogen atoms are replaced with one or more hydroxy groups.

[0118] "Thiol" or "sulfanyl" refers to –SH. Substituted thiol or sulfanyl refers to –S-R^{η}, where R^{η} is an alkyl, aryl, or other suitable substituent.

[0119] "Sulfonyl" refers to –SO_2-. Substituted sulfonyl refers to –SO_2-R^{η}, where R^{η} is an alkyl, aryl, or other suitable substituent.

[0120] "Alkylsulfonyl" refers to –SO_2-R^{ζ}, where R^{ζ} is an alkyl, which can be optionally substituted. Typical alkylsulfonyl groups include, but are not limited to, methylsulfonyl, ethylsulfonyl, n-propylsulfonyl, and the like.

[0121] "Phosphate" as used herein refers to a functional group comprised of an orthophosphate ion (phosphorous atom covalently linked to four oxygen atoms). The orthophosphate ion is commonly found with one or more hydrogen atoms or organic groups.

[0122] "Phosphorylated" as used herein refers to the addition or presence of one of more phosphoryl groups (phosphorous atom covalently linked to the three oxygen atoms).

[0123] "Optionally substituted" as used herein with respect to the foregoing chemical groups means that positions of the chemical group occupied by hydrogen can be substituted with another atom (unless otherwise specified) exemplified by, but not limited to carbon, oxygen, nitrogen, or sulfur, or a chemical group, exemplified by, but not limited to, hydroxy, oxo, nitro, methoxy, ethoxy, alkoxy, substituted alkoxy, trifluoromethoxy, haloalkoxy, fluoro, chloro, bromo, iodo, halo, methyl, ethyl, propyl, butyl, alkyl, alkenyl, alkynyl, substituted alkyl, trifluoromethyl, haloalkyl, hydroxyalkyl, alkoxyalkyl, thio, alkylthio, acyl, carboxy, alkoxycarbonyl, carboxamido, substituted carboxamido, alkylsulfonyl, alkylsulfinyl, alkylsulfonylamino, sulfonamido, substituted sulfonamido, cyano, amino, substituted amino, alkylamino, dialkylamino, aminoalkyl, acylamino, amidino, amidoximo, hydroxamoyl, phenyl, aryl, substituted aryl, aryloxy, arylalkyl, arylalkenyl, arylalkynyl, pyridyl, imidazolyl, heteroaryl, substituted heteroaryl, heteroaryloxy, heteroarylalkyl, heteroarylalkenyl, heteroarylalkynyl, cyclopropyl, cyclobutyl, cyclopentyl, cyclohexyl, cycloalkyl, cycloalkenyl, cycloalkylalkyl, substituted cycloalkyl, cycloalkyloxy, pyrrolidinyl, piperidinyl, morpholino, heterocycle, (heterocycle)oxy, and (heterocycle)alkyl; where preferred heteroatoms are oxygen, nitrogen, and sulfur. Additionally, where open valences exist on these substitute chemical groups they can be further substituted with alkyl, cycloalkyl, aryl, heteroaryl, and/or heterocycle groups, that where these open valences exist on carbon they can be further substituted by halogen and by oxygen-, nitrogen-, or sulfur-bonded substituents, and where multiple such open valences exist, these groups can be joined to form a ring, either by direct formation of a bond or by formation of bonds to a new heteroatom, preferably oxygen, nitrogen, or sulfur. It is further contemplated that the above substitutions can be made provided that replacing the hydrogen with the substituent does not introduce unacceptable instability to the molecules of the present disclosure and is otherwise chemically reasonable. One of ordinary skill in the art would understand that with respect to any

chemical group described as optionally substituted, only sterically practical and/or synthetically feasible chemical groups are meant to be included. "Optionally substituted" as used herein refers to all subsequent modifiers in a term or series of chemical groups. For example, in the term "optionally substituted arylalkyl," the "alkyl" portion and the "aryl" portion of the molecule may or may not be substituted, and for the series "optionally substituted alkyl, cycloalkyl, aryl and heteroaryl," the alkyl, cycloalkyl, aryl, and heteroaryl groups, independently of the others, may or may not be substituted.

[0124] "Reaction" as used herein refers to a process in which one or more substances or compounds or substrates is converted into one or more different substances, compounds, or processes.

Template-Independent Synthesis by Engineered TnTs

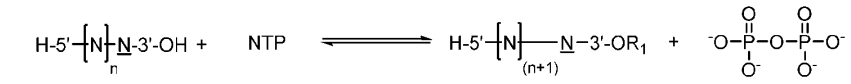
[0125] New methods of efficiently synthesizing high purity strands of DNA, RNA, and other polynucleotides are necessary to overcome the limitations of existing phosphoramidite chemical synthesis in order to enable a range of emerging and existing synthetic biology applications.

[0126] The present disclosure provides novel terminal nucleotidyl transferases that have improved activity in the template-independent synthesis of polynucleotides, including therapeutic RNAs, using 5'-nucleoside triphosphates ("NTPs") modified with a 3'-O-removable blocking group (NTP-3'-O-RBG) or other natural or modified NTP substrates. The TnTs of the present disclosure have improved thermostability, increased activity at elevated temperatures, increased soluble expression or isolated protein yield, decreased by-product formation, increased specific activity on NTP-3'-O-RBG or natural or modified NTP substrates, increased incorporation efficiency in extension of oligo acceptor substrates, and/or increased activity on various oligo acceptor substrates,

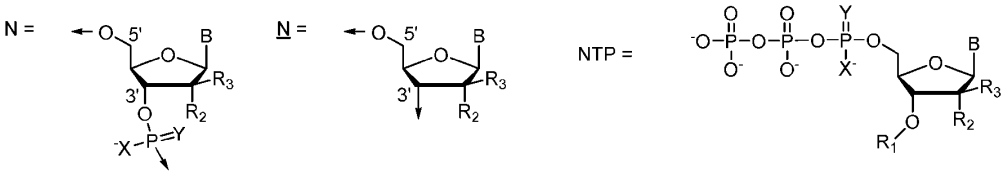
[0127] In particular, the TnTs of the present disclosure have increased specific activity on NTP-3'-O-RBGs or natural or modified NTP substrates comprising a conjugate moiety, reactive group, or linker; increased incorporation efficiency in extension of oligo acceptor substrates comprising a conjugate moiety, reactive group, or linker; and/or increased activity on various oligo acceptor substrates comprising a conjugate moiety, reactive group, or linker as compared to a wild-type TnT (or TdT) or other TnTs or template-independent polymerases known to those of skill in the art. The engineered polypeptides of the present disclosure are variants of the TdT enzyme of SEQ ID NO: 10, which is a predicted splice variant of the wild-type gene from *Monodelphis domestica*.

[0128] Template-independent synthesis of a defined polynucleotide sequence using an engineered TnT is a multistep process. In one embodiment, an oligo acceptor substrate with a 3'-OH allows addition of a defined modified NTP substrate (in this example, an NTP-3'-O-RBG) by an engineered TnT, as depicted in Scheme 1, below.

Scheme 1



where



and

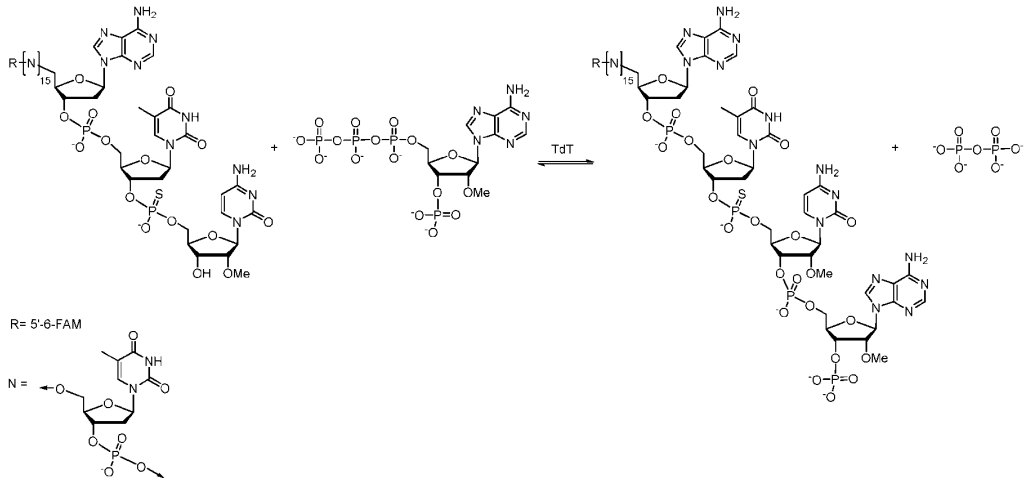
B = adenine, guanine, cytidine, thymine, uracil, pseudouridine, 1-methylpseudouridine

R_1 = benzyl, *O*-nitrolbenzyl, benzoyl, acetyl, cyanoethyl, NO_2, PO_3^{2-}, SO_3^-
R_2 = H, OH, OMe, F, Me, methoxyethyl (MOE)
R_3 = H, OH, OMe, F, Me
X = O,S
Y = O,S

[0129] After reaction of the NTP-3'-O-RBG with the 3'-OH of oligo acceptor substrate or the growing polynucleotide chain, the TnT is blocked from further reaction by the 3'-O-RBG. The RBG is then removed, exposing the 3'-OH and allowing another round of addition. After each round of addition, the blocking group of the nucleotide-3'-O-RBG or natural or modified nucleotide from the previous round is removed, and a new NTP-3'-O-RBG or natural or modified NTP substrate is added to sequentially and efficiently create a defined polynucleotide sequence by addition at the 3'-OH end of the polynucleotide or oligo acceptor substrate without a complimentary strand or templating primer sequence. After synthesis of the defined polynucleotide is complete, the oligonucleotide chain may be cleaved or released from the oligo acceptor substrate.

[0130] A variety of oligo acceptor substrates and NTP-3'-O-RBG or natural or modified NTP substrates may be used in this process, as may be envisioned by one of skill in the art. An example of one reaction is detailed in Scheme 2, below. Scheme 2 depicts the TnT catalyzed reaction of 5'-6-FAM-$[N]_{15}$AT*mC and 3'-phos-mATP, while other examples of suitable oligo acceptor substrate and NTP-3'-O-RBG or natural or modified NTP pairs are described in the Examples. These examples are non-limiting.

Scheme 2



[0131] Occasionally, undesired synthesis products are created by the TnT during the addition step. This includes incorporation of NTPs that have lost their blocking group, addition of more than one NTP, or the excision or pyrophosphorolysis of the TnT on the growing polynucleotide chain.

[0132] In some embodiments, one or more additional quality control steps are used, such as adding an exonuclease prior to removing the blocking group and initiating a new round of synthesis. In some embodiments, a phosphatase, such as a pyrophosphatase, is used to breakdown inorganic phosphate and push the reversible TnT reaction toward synthesis.

[0133] As described further herein, the engineered TnT polypeptides of the current disclosure exhibit one of more improved properties in the template-independent polynucleotide synthesis process depicted in Schemes 1 and 2.

[0134] In some embodiments, the present disclosure provides an engineered TnT polypeptide comprising an amino acid sequence having at least 60% sequence identity to an amino acid reference sequence of SEQ ID NO: 10 and further comprising one or more amino acid residue differences as compared to the reference amino acid sequence, wherein the engineered TnT polypeptide has improved thermostability, increased activity at elevated temperatures, increased soluble expression or isolated protein yield, decreased by-product formation, increased specific activity on NTP-3'-O-RBG or natural or modified NTP substrates increased incorporation efficiency in extension of oligo acceptor substrates, and/or increased activity on various oligo acceptor substrates,

[0135] In particular, the TnTs of the present disclosure have increased specific activity on NTP-3'-O-RBGs or natural or modified NTP substrates comprising a conjugate moiety, reactive group, or linker; increased incorporation efficiency in extension of oligo acceptor substrates comprising a conjugate moiety, reactive group, or linker; and/or increased activity on various oligo acceptor substrates

comprising a conjugate moiety, reactive group, or linker as compared to a wild-type TnT (or TdT) or other TnTs or template-independent polymerases known to those of skill in the art.

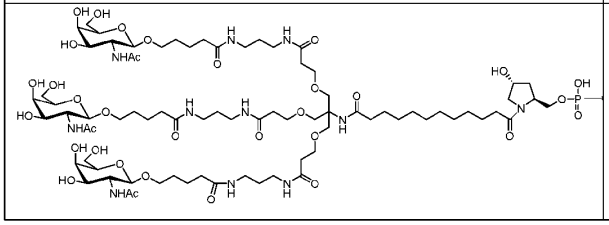
Use of Conjugate Moieties, Reactive Groups, and Linkers in RNA Therapeutics

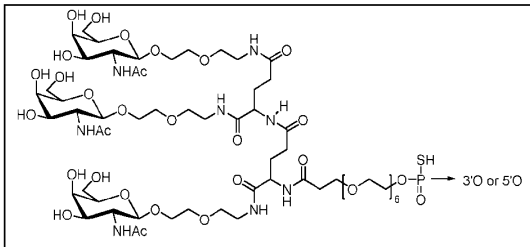
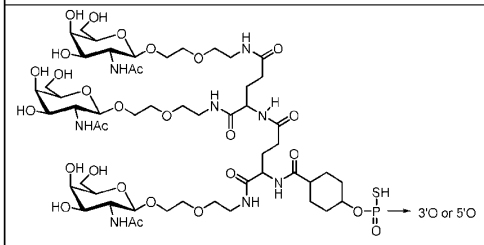
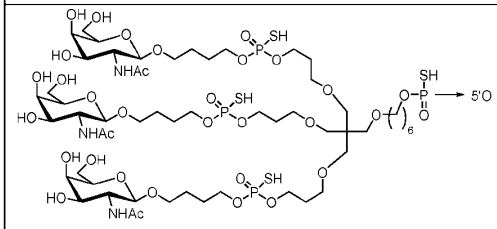
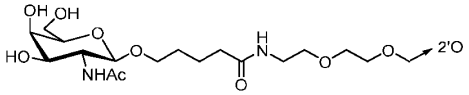
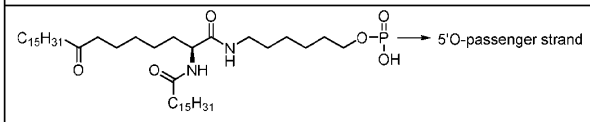
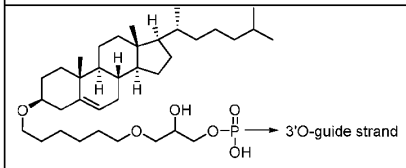
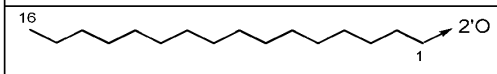
[0136] The therapeutic potential of antisense oligonucleotides (ASO), which act to silence target mRNA via RNAse H degradation of the duplex ASO bound to target mRNA, and of siRNAs, which act primarily through RISC complex binding to silence mRNA translation and/or transcription, has long been recognized. Experience with the design and implementation of first generation ASO and siRNA therapeutics has emphasized the importance of nucleic acid and nucleotide modification to enhance the efficacy of these molecules.

[0137] Incorporation of modified nucleotides (e.g. nucleotides with an additional chemical group at the 2' position of the sugar moiety) has been widely adopted to improve tissue targeting, potency, and stability of therapeutic oligonucleotides, while maintaining an A-form helix and compatibility with the RISC complex. For example, the strategic placement of 2′-fluoro and 2′-O-methyl modifications is known to increase potency. Similarly, a design incorporating a phosphate analog, such a vinyl phosphate, at the 5' nucleotide of the antisense strand may increase siRNA potency by enhancing RISC complex binding.

[0138] While many nucleotide modifications function to prevent nuclease degradation and enhance stability in the cellular environment, the introduction of other nucleotide modifications for organ and tissue specific targeting has become increasingly prevalent. Various conjugate moieties have been utilized to target FDA-approved RNA therapeutics to eye, skeletal muscle, spinal cord, and liver tissues. Targeting conjugate moieties often comprise larger, more complex structures as compared to common 2'-fluoro and 2'-O-methyl modifications. As an example, lipid chains such as 2'-O-hexadecyl (C_{16}) and moieties comprising one or more GalNAc (N-acetylgalactosamine) molecules are used to target various siRNA therapeutics that are currently FDA-approved or under development (see Table 2, below).

Table 2.

Structure	Location on strand	Compound
	3'O of passenger strand	L96 (Gal-NAc_3 ligand)

	5'O of guide strand	conjugation in Olpasiran
	3'O of passenger strand	conjugation in Plozasiran and Fazirsiran and Zodasiran
	5'O of guide strand	Zerlasiran
	2'O conjugation, not at termini	compound can have 4 nucleotides in a row on hairpin
	5'O of passenger strand	DXT
	3'O of guide strand	conjugation in Remlarsen
	2'O modification at around the 6th nucleotide	C_{16} or 2'-O-hexadecyl

[0139] In addition to the functions discussed above, conjugate moieties may have other uses in polynucleotides in RNA therapeutics or in other contexts. An example of this is the use of cleavable linkers, as discussed above in the Introduction, which may be useful as blocking groups for

polynucleotide synthesis or may have other functions, such as photosensing, labeling, or other sensing reactions.

[0140] Many examples of useful conjugate moieties are known in the art. These include saccharide, lectin, lipid, sterol, peptide, protein, antibody, aptamer, peptide nucleic acid, vitamin, toxin, hormone, and other molecules, as described further herein. (Roberts et al. Nat Rev Drug Discov 19, 673–694 (2020), Shi et al. Chem Rev. 2024;124(3):929-1033,US 10,294,474, US20220170025A1, Nair et al. J Am Chem Soc. 2014;136(49):16958-16961, Kumar & Turnbull. Chem Soc Rev. 2023;52(4):1273-1287).

[0141] In some examples, a conjugate moiety may comprise a reactive group that provides a site for further chemical modification or synthesis. Reactive groups are well known and include azido, amino, cyano, alkynyl groups, and activated esters and amines, among others. These groups could be incorporated into a polynucleotide chain to enable further useful conjugation or synthetic reactions in a subsequent step or process. One example familiar to those in the art is the use of click chemistry, which has been widely used to introduce modifications into nucleic acids (Antoni et al. Chem Rev. 2021;121(12):7122-7154).

[0142] Conjugate moieties or reactive groups may be attached by a linker, such as a glycol or alkyl linker. Other linkers can be envisaged by one of skill in the art. Linkers may be used to link any position of the NTP or nucleotide to another moiety or functional group or may be incorporated for further functionalization or linking to another moiety or group at a later time.

Location of Conjugate Moieties, Reactive Groups, and Linkers

[0143] As depicted in Table 2, NTP-3'-O-RBGs or natural or modified NTPs comprising conjugate moieties, reactive groups, and linkers may be located at a variety of positions in a polynucleotide, including positions in both the passenger and/or the guide strands of siRNAs.

[0144] The NTP-3'-O-RBG or natural or modified NTP comprising the conjugate moiety, reactive group, or linker may be at the 3'-O terminal position, as an internal modification (including at the hairpin turn), or, although uncommon, at the 5'-O terminal position of either strand. Conjugate moieties, reactive groups, or linkers present as internal modifications may be at any nucleotide position, although location at specific nucleotide positions may enhance the function or activity of certain conjugate moieties, reactive groups, and linkers.

[0145] The RNA or other polynucleotide may comprise one or more conjugate moieties, reactive groups, or linkers at one or more positions. In some embodiments, a RNA or other polynucleotide may comprise a nucleotide comprising a linker and a conjugate moiety or a nucleotide comprising a linker and a reactive group. In some embodiments, a RNA or other polynucleotide comprising more than one conjugate moieties, reactive groups, or linkers at more than one position may comprise the

same conjugate moiety, reactive group, or linker (or combinations thereof) at more than one position. In some embodiments, a RNA or other polynucleotide comprising more than one conjugate moieties, reactive groups, or linkers at more than one position may comprise a different conjugate moiety, reactive group, or linker (or combinations thereof) at more than one position.

Enzymatic Approaches to Incorporating Conjugate Moieties and Chemical Reactive Groups

[0146] As enzymatic approaches to synthesis of RNA therapeutics and other polynucleotides becomes feasible, strategies for incorporation of conjugate moieties, reactive groups, and linkers become necessary. Several enzymatic approaches utilizing template independent polymerases, such as a TnT of the present disclosure, or nucleotide ligases or both polymerases and ligases can be envisaged.

[0147] In some approaches, the conjugate may be chemically or enzymatically appended to an initiator oligonucleotide, which is then extended through action of a TnT. Similarly, a TnT may be used to synthesize part of a polynucleotide, which may then be ligated to a second polynucleotide containing the conjugate moiety, reactive group, or linker. A third enzyme may also be used in some synthesis methods.

[0148] In some other cases, a TnT or ligase may be capable of catalyzing the direct addition of the NTP-3'-O-RBG or natural or modified NTP comprising a conjugate moiety, reactive group, or linker to a polynucleotide chain or oligo acceptor substrate. In some embodiments, the TnT or ligase may catalyze the direct addition of the NTP-3'-O-RBG or natural or modified NTP comprising a conjugate moiety or reactive group to a polynucleotide chain or oligo acceptor substrate so that the NTP-3'-O-RBG or natural or modified NTP comprising the conjugate moiety, reactive group, or linker is at the 5' terminal position or is at the 3' terminal position or is at an internal position in the polynucleotide chain or oligo acceptor substrate.

[0149] In some other embodiments, both a ligase and a TnT may be used to add more than one conjugate moieties, reactive groups, or linkers at more than one location in a polynucleotide chain.

[0150] In some embodiments, a ligase and/or a TnT may be used to add a NTP-3'-O-RBG or natural or modified NTP comprising a conjugate moiety, reactive group, or linker to one or more positions in the polynucleotide chain or oligo acceptor substrate, wherein the conjugate moiety, reactive group, or linker may be used in a second step. In some embodiments, the second step is a further chemical or biocatalytic modification at the site of the conjugate moiety, reactive group, or linker.

[0151] As described herein and in a previous disclosure (PCT/US2023/076667), terminal nucleotidyl transferases with activity in the incorporation of backbone phosphorothiate links and NTPs comprising various small 2' modifications (e.g. fluoro, O-methyl, and O-methoxyethyl) into an oligo acceptor substrate or growing RNA chain have been identified.

[0152] However, polymerases and template independent polymerases, such as terminal nucleotidyl transferases, with activity incorporating NTP-3'-O-RBGs or natural or modified NTPs comprising conjugate moieties, reactive groups, or linkers are not known. Similarly, polymerases and template independent polymerases, such as terminal nucleotidyl transferases, with activity adding additional NTP-3'-O-RBGs or natural or modified NTP substrates onto a polynucleotide or oligo acceptor substrate containing one or more conjugate moieties, reactive groups, or linkers have not been identified.

[0153] Therefore, the present disclosure provides terminal nucleotidyl transferases with activity incorporating NTP-3'-O-RBGs or natural or modified NTPs comprising conjugate moieties, reactive groups, or linkers onto a polynucleotide or oligo acceptor substrate and capable of adding additional NTP-3'-O-RBGs or natural or modified NTP substrates onto a polynucleotide or oligo acceptor substrate containing one or more conjugate moieties, reactive groups, or linkers, as detailed in the Examples, below.

Engineered Terminal Nucleotidyl Transferase (TnT) Polypeptides

[0154] The present disclosure provides engineered terminal nucleotidyl transferase (TnT) polypeptides useful in template-independent polynucleotide synthesis using an NTP-3'-O-RBG or natural or modified NTP substrate, as well as compositions and methods of utilizing these engineered polypeptides in template-independent oligonucleotide synthesis.

[0155] In particular, the present disclosure provides TnT polypeptides with activity incorporating NTPs comprising conjugate moieties, reactive groups, or linkers onto a polynucleotide or oligo acceptor substrate and capable of adding additional NTP-3'-O-RBGs or natural or modified NTP substrates onto a polynucleotide or oligo acceptor substrate comprising one or more conjugate moieties, reactive groups, or linkers.

[0156] The present disclosure also provides TnT polypeptides, polynucleotides encoding the polypeptides, methods of preparing the polypeptides, and methods for using the polypeptides. Where the description relates to polypeptides, it is to be understood that it can describe the polynucleotides encoding the polypeptides.

[0157] Suitable reaction conditions under which the above-described improved properties of the engineered polypeptides carry out the desired reaction can be determined with respect to concentrations or amounts of polypeptide, substrate, co-substrate, buffer, solvent, pH, conditions including temperature and reaction time, and/or conditions with the polypeptide immobilized on a solid support, as further described below and in the Examples.

[0158] In some embodiments, exemplary engineered TnTs comprise an amino acid sequence that has one or more residue differences as compared to SEQ ID NO: 4, 6, 8, 10, and/or 12.

[0159] The structure and function information for the exemplary engineered polypeptides of the present disclosure are based on the conversion of an oligo acceptor substrate and a NTP -3'-O-RBG or a dideoxy NTP (e.g. a 2',3'-dideoxy NTP) comprising a conjugate moiety, reactive group, or linker, the results of which are shown below in the Examples.

[0160] The odd numbered sequence identifiers (i.e., SEQ ID NOs) in this disclosure refer to the nucleotide sequence encoding the amino acid sequence provided by the even numbered SEQ ID NOs in this disclosure. Exemplary sequences are provided in the electronic sequence listing file accompanying this invention, which is hereby incorporated by reference herein. The amino acid residue differences are based on comparison to the reference sequence of SEQ ID NOs: 4, 6, 8, 10, and/or 12.

[0161] Terminal deoxynucleotidyl transferase, a member of the Pol X family, has been identified in many species. Members of the diverse Pol X family are known to share certain residues, which are conserved across family members. TdT also has a high level of conservation across species for residues thought to be involved in binding divalent metal ions, ternary complex formation, and binding dNTP and DNA ligands (Domínguez et al. (2000). EMBO, 19(7), 1731–1742.) Additionally, TdTs are known to have splice variants which are N-terminal truncations, lacking a BRCT domain. Other template-independent polymerases (including, but not limited to polyA polymerases, polyU polymerases and terminal urildylytransferases) are also known in the art and may be used to practice the invention. Similarly, other polymerases are known to be capable of template-independent synthesis (including but not limited to reverse transcriptases) and may be used to practice the invention. As used herein, the term terminal nucleotidyl transferase (TnT) is used to distinguish an engineered enzyme with activity on a variety of nucleoside triphosphates, including ribonucleoside triphosphates with 3' modifications or with 2' modifications or both, from a wild-type TdT enzyme having wild-type TdT activity.

[0162] Various engineered or variant TnTs and template independent polymerases are known in the art , including terminal nucleotidyl transferases (TdTs), Pol X polymerases, Poly(N) polymerases, Polμ polymerases, Polβ polymerases, Polλ polymerases, and Polθ polymerases, and others (see for example, U.S. Pat. 10059929, U.S. Pat. 10,760,063, U.S. Pat. 10,774,316, U.S. Pat. 11,390,858, U.S. Pat. 10,745,727, PCT/GB2020/050247, US20210164008A1, US20210407509, WO2016GB50301, WO17216472, WO18215803, WO20072715, WO20077227, WO21122539, WO2022029427, and WO21116270, from Molecular Assemblies, Nuclera, DNA Script, and others, each of which is specifically incorporated herein).

[0163] The TdT enzyme of SEQ ID NO: 10 is a predicted splice variant of the wild-type gene from *Monodelphis domestica.* The TnT polypeptides of the present disclosure (SEQ ID NO: 4, 6, 8, and 12) are engineered variants of SEQ ID NO: 10 with a N-terminal 6-histidine tag.

[0164] The polypeptides of the present disclosure have residue differences that result in improved properties necessary to develop an efficient TnT enzyme, capable of template-independent synthesis of polynucleotides having a defined sequence. Various residue differences, at both conserved and non-conserved positions, have been discovered to be related to improvements in various enzymes properties, including improved thermostability, increased activity at elevated temperatures, increased soluble expression or isolated protein yield, decreased by-product formation, increased specific activity on NTP-3'-O-RBG or natural or modified NTP substrates, increased incorporation efficiency in extension of oligo acceptor substrates, increased activity on various oligo acceptor substrates, increased specific activity on NTP-3'-O-RBG or natural or modified NTP substrates comprising a conjugate moiety, reactive group, or linker, increased incorporation efficiency in extension of oligo acceptor substrates comprising a conjugate moiety, reactive group, or linker, and/or increased activity on various oligo acceptor substrates comprising a conjugate moiety, reactive group, or linker as compared to a wild-type TnT (or TdT) or other TnTs or template-independent polymerases known to those of skill in the art. In some embodiments, the engineered TnT polypeptides exhibit increased incorporation efficiency of greater than 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% in extension of an oligo acceptor substrate by addition of an NTP or NQP comprising a conjugate moiety, reactive group, or linker.

[0165] In light of the guidance provided herein, it is further contemplated that any of the exemplary engineered polypeptides comprising the sequences of SEQ ID NOs: 4, 6, 8, and 12 find use as the starting amino acid sequence for synthesizing other TnT polypeptides. Further improvements may be generated by including amino acid differences at residue positions that had been maintained as unchanged throughout earlier rounds of evolution.

[0166] In some embodiments, the engineered TnT polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: **10** and one or more residue differences as compared to SEQ ID NO: **10** at a position or set of positions selected from 12; 14; 26; 30; 34; 37; 38; 53; 79; 80; 81; 90; 92; 94; 101; 108; 121; 137; 140; 141; 142; 150; 152; 153; 155; 160; 163; 165; 174; 177; 184; 185; 189; 190; 194; 196; 201; 203; 205; 213; 219; 231; 244; 248; 258; 263; 264; 273; 284; 288; 289; 290; 293; 300; 304; 307; 313; 314; 315; 317; 318; 324; 325; 333; 336; 340; 342; 344; 352; 353; 359; 362; 379; 390; 392; 393; 394; 397; 398; 401; 402; 403; 406; 408; 410; 411; 413; 414; 415; 416; 425; 427; 428; 429; 431; 433; 434; 441; 442; 444; 446; 455; 456; 460; 461; 462; 466; 468; 470; 474; 476; 481; 484; 485; 488; 495; 499; 501; 502; 506; 515; 522; 523; and 525. In some embodiments, the engineered TnT polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: **10** and one or more residue differences as compared to SEQ ID NO: **10** selected from 12S; 14G; 26Q; 30G; 34D; 37A;

38T; 53E; 79T; 80D; 81E; 90L; 92R; 94E; 101E; 108K; 121S; 137A; 140V; 141E; 142V; 150E; 152R; 153V; 155E; 160M; 163V; 165P; 174L; 177R; 184S; 185L; 189R; 190E; 194L; 196G; 201A; 203E; 205R; 213S; 219R; 231S; 244V; 248R; 258W; 263H; 264A; 273V; 284L; 288E; 289D; 290V; 293S; 300P; 304V; 307L; 313A; 314R; 315G; 317T; 318R; 324I; 325W; 333A; 336D; 340I; 342E; 344V; 352P; 353Q; 359L; 362S; 379M; 390; 392G; 393R; 394L; 397Q; 398W; 401G; 402G; 403S; 406G; 408A; 410E; 411G; 413A; 414E; 415A; 416S; 425D; 427L; 428V; 429R; 431S; 433R; 434N; 441M; 442G; 444S; 446P; 455L; 456R; 460G; 461S; 462F; 466M; 468Q; 470T; 474M; 476R; 481W; 484L; 485E; 488N; 495S; 499M; 501R; 502G; 506P; 515E; 522L; 523E; and 525F. In some embodiments, the engineered TnT polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: **10** and one or more residue differences as compared to SEQ ID NO: **10** selected from M12S; R14G; R26Q; M30G; I34D; M37A; I38T; K53E; S79T; N80D; S81E; N90L; G92R; D94E; T101E; T108K; C121S; M137A; R140V; V141E; D142V; A150E; T152R; L153V; I155E; T160M; I163V; Q165P; I174L; H177R; A184S; F185L; A189R; K190E; F194L; E196G; C201A; T203E; M205R; C213S; V219R; G231S; D244V; L248R; R258W; K263H; L264A; L273V; F284L; N288E; K289D; I290V; D293S; K300P; A304V; C307L; I313A; D314R; C315G; S317T; K318R; V324I; S325W; W333A; L336D; L340I; T342E; T344V; E352P; F353Q; F359L; T362S; T379M; Y390; D392G; L393R; I394L; T397Q; F398W; L401G; K402G; L403S; R406G; I408A; A410E; L411G; H413A; F414E; Q415A; K416S; H425D; K427L; E428V; D429R; R431S; W433R; E434N; E441M; S442G; A444S; S446P; V455L; V456R; D460G; R461S; Y462F; L466M; G468Q; S470T; Q474M; E476R; R481W; T484L; H485E; K488N; A495S; K499M; K501R; K502G; K506P; A515E; I522L; Q523E; and S525F.

[0167] In some embodiments, the engineered TnT polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: **10** and one or more residue differences as compared to SEQ ID NO: **10** at a position or set of positions selected from 12; 14; 26; 30; 34; 37; 38; 53; 79; 80; 81; 90; 92; 94; 101; 108; 121; 137; 140; 141; 142; 150; 152; 153; 155; 160; 163; 165; 174; 177; 184; 185; 189; 190; 193; 196; 201; 203; 205; 213; 219; 231; 244; 248; 258; 263; 264; 267; 273; 274; 281; 284; 288; 289; 290; 293; 300; 301; 302; 304; 307; 313; 314; 315; 317; 318; 324; 325; 333; 336; 340; 342; 344; 347; 352; 353; 354; 359; 362; 379; 390; 391; 392; 393; 394; 395; 397; 398; 400; 401; 402; 406; 408; 410; 411; 413; 414; 415; 416; 425; 427; 428; 429; 431; 433; 434; 441; 442; 444; 446; 451; 455; 456; 459; 460; 461; 462; 466; 468; 470; 474; 476; 481; 484; 485; 488; 495; 499; 501; 502; 504; 506; 515; 522; 523; 525; and 526. In some embodiments, the engineered TnT polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: **10** and one or more residue differences as compared to

SEQ ID NO: **10** selected from 12S; 14G; 26Q; 30G; 34D; 37A; 38T; 53E; 79T; 80D; 81E; 90L; 92R; 94E; 101E; 108K; 121S; 137A; 140V; 141E; 142V; 150E; 152R; 153V; 155E; 160M; 163V; 165P; 174L; 177R; 184S; 185L; 189R; 190E; 193V; 196G; 201A; 203E; 205R; 213S; 219R; 231S; 244V; 248R; 258W; 263H; 264A; 267R; 273V; 274G; 281Q; 284L; 288E; 289D; 290V; 293S; 300T; 301W; 302A; 304V; 307T; 313A; 314R; 315G; 317T; 318R; 324I; 325W; 333A; 336D; 340I; 342E; 344V; 347Q; 352P; 353Q; 354S; 359L; 362S; 379M; 390C; 391G; 392K; 393V; 394Y; 395W; 397Q; 398W; 400W; 401G; 402G; 406G; 408A; 410E; 411G; 413G; 414E; 415A; 416S; 425D; 427L; 428V; 429R; 431S; 433H; 434N; 441M; 442G; 444S; 446P; 451K; 455L; 456R; 459R; 460G; 461S; 462F; 466M; 468Q; 470T; 474M; 476R; 481D; 484R; 485E; 488N; 495S; 499M; 501R; 502G; 504Q; 506P; 515E; 522L; 523E; 525F; and 526Y. In some embodiments, the engineered TnT polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: **10** and one or more residue differences as compared to SEQ ID NO: **10** selected from M12S; R14G; R26Q; M30G; I34D; M37A; I38T; K53E; S79T; N80D; S81E; N90L; G92R; D94E; T101E; T108K; C121S; M137A; R140V; V141E; D142V; A150E; T152R; L153V; I155E; T160M; I163V; Q165P; I174L; H177R; A184S; F185L; A189R; K190E; E193V; E196G; C201A; T203E; M205R; C213S; V219R; G231S; D244V; L248R; R258W; K263H; L264A; S267R; L273V; K274G; R281Q; F284L; N288E; K289D; I290V; D293S; K300T; M301W; Q302A; A304V; C307T; I313A; D314R; C315G; S317T; K318R; V324I; S325W; W333A; L336D; L340I; T342E; T344V; F347Q; E352P; F353Q; G354S; F359L; T362S; T379M; Y390C; C391G; D392K; L393V; I394Y; E395W; T397Q; F398W; D400W; L401G; K402G; R406G; I408A; A410E; L411G; H413G; F414E; Q415A; K416S; H425D; K427L; E428V; D429R; R431S; W433H; E434N; E441M; S442G; A444S; S446P; R451K; V455L; V456R; Y459R; D460G; R461S; Y462F; L466M; G468Q; S470T; Q474M; E476R; R481D; T484R; H485E; K488N; A495S; K499M; K501R; K502G; F504Q; K506P; A515E; I522L; Q523E; S525F; and E526Y.

[0168] In some embodiments, the engineered TnT polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: **10** and one or more residue differences as compared to SEQ ID NO: **10** at a position or set of positions selected from 12; 14; 26; 30; 34; 37; 38; 53; 79; 80; 81; 90; 92; 94; 101; 108; 121; 137; 140; 141; 142; 150; 152; 153; 155; 160; 163; 165; 174; 177; 184; 185; 189; 190; 193; 196; 200; 201; 203; 205; 213; 219; 231; 244; 248; 258; 263; 264; 273; 274; 277; 281; 284; 288; 289; 290; 293; 300; 301; 302; 304; 307; 313; 314; 315; 317; 318; 324; 325; 333; 336; 340; 342; 344; 347; 352; 353; 354; 359; 362; 379; 380; 381; 390; 391; 392; 393; 394; 395; 397; 398; 400; 401; 402; 404; 406; 408; 410; 411; 413; 414; 415; 416; 425; 427; 428; 429; 431; 433; 434; 441; 442; 444; 446; 455; 456; 459; 460; 461; 462; 466; 468; 470; 474; 476; 477; 481; 484; 485; 488; 492; 495; 499; 501; 502; 504; 506; 515; 522; 523; 525; and 526. In

some embodiments, the engineered TnT polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: **10** and one or more residue differences as compared to SEQ ID NO: **10** selected from 12S; 14G; 26Q; 30G; 34D; 37A; 38T; 53E; 79T; 80D; 81E; 90L; 92R; 94E; 101E; 108K; 121S; 137A; 140V; 141E; 142V; 150E; 152R; 153V; 155E; 160M; 163V; 165P; 174L; 177R; 184S; 185L; 189R; 190E; 193V; 196G; 200K; 201A; 203E; 205R; 213S; 219R; 231S; 244V; 248R; 258W; 263H; 264A; 273R; 274P; 277S; 281Q; 284L; 288E; 289D; 290V; 293S; 296V; 300T; 301W; 302A; 304V; 307T; 313A; 314R; 315G; 317T; 318R; 324I; 325W; 333A; 336D; 340I; 342E; 344V; 347Q; 352P; 353Q; 354S; 359L; 362S; 379M; 380R; 381Y; 390C; 391G; 392K; 393V; 394Y; 395W; 397Q; 398W; 400W; 401G; 402G; 404S; 406G; 408A; 410E; 411G; 413G; 414E; 415A; 416S; 425D; 427L; 428V; 429R; 431S; 433H; 434N; 441M; 442G; 444S; 446P; 455L; 456R; 459R; 460G; 461S; 462M; 466M; 468Q; 470T; 474M; 476R; 477Q; 481D; 484R; 485E; 488N; 492M; 495S; 499M; 501R; 502G; 504Q; 506P; 515E; 522L; 523E; 525F; and 526Y. In some embodiments, the engineered TnT polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: **10** and one or more residue differences as compared to SEQ ID NO: **10** selected from M12S; R14G; R26Q; M30G; I34D; M37A; I38T; K53E; S79T; N80D; S81E; N90L; G92R; D94E; T101E; T108K; C121S; M137A; R140V; V141E; D142V; A150E; T152R; L153V; I155E; T160M; I163V; Q165P; I174L; H177R; A184S; F185L; A189R; K190E; E193V; E196G; T200K; C201A; T203E; M205R; C213S; V219R; G231S; D244V; L248R; R258W; K263H; L264A; L273R; K274P; D277S; R281Q; F284L; N288E; K289D; I290V; D293S; L296V; K300T; M301W; Q302A; A304V; C307T; I313A; D314R; C315G; S317T; K318R; V324I; S325W; W333A; L336D; L340I; T342E; T344V; F347Q; E352P; F353Q; G354S; F359L; T362S; T379M; N380R; L381Y; Y390C; C391G; D392K; L393V; I394Y; E395W; T397Q; F398W; D400W; L401G; K402G; P404S; R406G; I408A; A410E; L411G; H413G; F414E; Q415A; K416S; H425D; K427L; E428V; D429R; R431S; W433H; E434N; E441M; S442G; A444S; S446P; V455L; V456R; Y459R; D460G; R461S; Y462M; L466M; G468Q; S470T; Q474M; E476R; R477Q; R481D; T484R; H485E; K488N; D492M; A495S; K499M; K501R; K502G; F504Q; K506P; A515E; I522L; Q523E; S525F; and E526Y.

[0169] In some embodiments, the engineered TnT polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: **10** and one or more residue differences as compared to SEQ ID NO: **10** at a position or set of positions selected from 12; 14; 26; 30; 34; 37; 38; 53; 79; 80; 81; 90; 92; 94; 101; 108; 121; 137; 140; 141; 142; 150; 152; 153; 155; 160; 163; 165; 174; 177; 184; 185; 189; 190; 193; 196; 200; 201; 203; 205; 208; 213; 219; 231; 244; 248; 258; 263; 264; 273; 274; 277; 281; 284; 288; 289; 290; 293; 296; 299; 300; 301; 302; 304;

307; 313; 314; 315; 317; 318; 319; 324; 325; 333; 336; 340; 342; 343; 344; 347; 352; 353; 354; 359; 362; 363; 379; 380; 381; 390; 391; 392; 393; 394; 395; 397; 398; 400; 401; 402; 404; 406; 407; 408; 410; 411; 413; 414; 415; 416; 425; 427; 428; 429; 431; 433; 434; 441; 442; 444; 446; 455; 456; 459; 460; 461; 462; 465; 466; 468; 470; 474; 476; 477; 481; 484; 485; 488; 492; 495; 499; 501; 502; 504; 506; 515; 522; 523; 525; and 526. In some embodiments, the engineered TnT polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: **10** and one or more residue differences as compared to SEQ ID NO: **10** selected from 12S; 14G; 26Q; 30G; 34D; 37A; 38T; 53E; 79T; 80D; 81E; 90L; 92R; 94E; 101E; 108K; 121S; 137A; 140V; 141E; 142V; 150E; 152R; 153V; 155E; 160M; 163V; 165P; 174L; 177R; 184S; 185L; 189R; 190R; 193S; 196G; 200K; 201A; 203E; 205R; 208E; 213S; 219R; 231S; 244V; 248R; 258W; 263H; 264A; 273R; 274P; 277S; 281Q; 284L; 288E; 289D; 290V; 293S; 296V; 299S; 300T; 301W; 302A; 304V; 307T; 313A; 314R; 315G; 317T; 318R; 319T; 324I; 325W; 333A; 336D; 340I; 342E; 343L; 344V; 347Q; 352P; 353Q; 354S; 359L; 362S; 363P; 379M; 380R; 381Y; 390C; 391G; 392K; 393V; 394Y; 395W; 397K; 398W; 400W; 401G; 402S; 404S; 406G; 407P; 408A; 410E; 411G; 413G; 414E; 415A; 416A; 425D; 427L; 428V; 429R; 431S; 433H; 434N; 441M; 442G; 444S; 446P; 455L; 456R; 459R; 460G; 461S; 462M; 465G; 466M; 468Q; 470T; 474M; 476R; 477Q; 481D; 484R; 485E; 488N; 492H; 495G; 499M; 501R; 502G; 504Q; 506P; 515E; 522L; 523E; 525F; and 526Y. In some embodiments, the engineered TnT polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: **10** and one or more residue differences as compared to SEQ ID NO: **10** selected from M12S; R14G; R26Q; M30G; I34D; M37A; I38T; K53E; S79T; N80D; S81E; N90L; G92R; D94E; T101E; T108K; C121S; M137A; R140V; V141E; D142V; A150E; T152R; L153V; I155E; T160M; I163V; Q165P; I174L; H177R; A184S; F185L; A189R; K190R; E193S; E196G; T200K; C201A; T203E; M205R; I208E; C213S; V219R; G231S; D244V; L248R; R258W; K263H; L264A; L273R; K274P; D277S; R281Q; F284L; N288E; K289D; I290V; D293S; L296V; T299S; K300T; M301W; Q302A; A304V; C307T; I313A; D314R; C315G; S317T; K318R; A319T; V324I; S325W; W333A; L336D; L340I; T342E; I343L; T344V; F347Q; E352P; F353Q; G354S; F359L; T362S; S363P; T379M; N380R; L381Y; Y390C; C391G; D392K; L393V; I394Y; E395W; T397K; F398W; D400W; L401G; K402S; P404S; R406G; K407P; I408A; A410E; L411G; H413G; F414E; Q415A; K416A; H425D; K427L; E428V; D429R; R431S; W433H; E434N; E441M; S442G; A444S; S446P; V455L; V456R; Y459R; D460G; R461S; Y462M; A465G; L466M; G468Q; S470T; Q474M; E476R; R477Q; R481D; T484R; H485E; K488N; D492H; A495G; K499M; K501R; K502G; F504Q; K506P; A515E; I522L; Q523E; S525F; and E526Y.

[0170] As will be appreciated by the skilled artisan, in some embodiments, one or a combination of residue differences above that is selected can be kept constant (i.e., maintained) in the engineered TnT

as a core feature, and additional residue differences at other residue positions incorporated into the sequence to generate additional engineered TnT polypeptides with improved properties. Accordingly, it is to be understood for any engineered TnT containing one or a subset of the residue differences above, the present disclosure contemplates other engineered TnTs that comprise the one or subset of the residue differences, and additionally one or more residue differences at the other residue positions disclosed herein.

[0171] As noted above, the engineered TnT polypeptides are also capable of converting substrates (e.g., NTP-3'-O-RBG comprising a conjugate moiety, reactive group, or linker and/or a natural or modified NTP and an oligo acceptor substrate comprising a conjugate moiety, reactive group, and/or linker) to products (e.g., an oligo acceptor substrate with an added nucleotide-3'-O-RBG). In some embodiments, the engineered TnT polypeptide is capable of converting the substrate compounds to the product compound with at least 1.2 fold, 1.5 fold, 2 fold, 3 fold, 4 fold, 5 fold, 10 fold, 20 fold, 30 fold, 40 fold, 50 fold, 60 fold, 70 fold, 80 fold, 90 fold, 100 fold, or more activity relative to the activity of the reference polypeptide of SEQ ID NOs: 4, 6, 8, 10, or 12. Exemplary substrates and products are provided in the Examples, below.

[0172] In some embodiments, the engineered TnT is capable of converting the substrate compounds to the product compounds with at least 2-fold the activity relative to SEQ ID NOs: 4, 6, 8, 10, or 12 or another TnT or template independent polymerase. In some embodiments, the engineered TnT is capable of converting the substrate compounds to the product compounds with at least 5-fold the activity relative to SEQ ID NOs: 4, 6, 8, 10, or 12 or another TnT or template independent polymerase. In some embodiments, the engineered TnT is capable of converting the substrate compounds to the product compounds with at least 20-fold the activity relative to SEQ ID NOs: 4, 6, 8, 10, or 12 or another TnT or template independent polymerase. In some embodiments, the engineered TnT is capable of converting the substrate compounds to the product compounds with at least 50-fold the activity relative to SEQ ID NOs: 4, 6, 8, 10, or 12 or another TnT or template independent polymerase.

[0173] In some embodiments, the engineered TnT has an amino acid sequence comprising one or more residue differences as compared to SEQ ID NOs: 4, 6, 8, 10, or 12, that reduces the by-product formation of the engineered TnT, as compared to a wild-type or engineered reference TnT.

[0174] In some embodiments, the engineered TnT has an amino acid sequence comprising one or more residue differences as compared to SEQ ID NOs: 4, 6, 8, 10, or 12, that increases specific activity of the engineered TnT on one or more NTP-3'-O-RBGs or natural or modified NTP substrates comprising a conjugate moiety, reactive group, or linker, as compared to a wild-type or engineered reference TnT.

[0175] In some embodiments, the engineered TnT has an amino acid sequence comprising one or more residue differences as compared to SEQ ID NOs: 4, 6, 8, 10, or 12, that increases specific activity of the engineered TnT on one or more oligo acceptor substrates comprising at least one conjugate moiety, reactive group, or linker, as compared to a wild-type or engineered reference TnT.

[0176] In some embodiments, the engineered TnT has an amino acid sequence comprising one or more residue differences as compared to SEQ ID NOs: 4, 6, 8, 10, or 12, that increases incorporation efficiency in extension of an oligo acceptor substrate by addition of an NTP or NQP substrate comprising a conjugate moiety, reactive group, or linker of greater than 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% when compared to the incorporation efficiency of a wild-type or engineered reference TnT.

[0177] In some embodiments, the engineered TnT with improved properties has an amino acid sequence comprising a sequence selected from the even-numbered sequences of SEQ ID NO: 4-8, and 12. In some embodiments, the engineered TnT with improved properties has an amino acid sequence comprising a sequence selected from SEQ ID NOs: 4, 6, 8, or 12.

[0178] In addition to the residue positions specified above, any of the engineered TnT polypeptides disclosed herein can further comprise other residue differences relative to SEQ ID NOs: 4, 6, 8, 10, or 12, at other residue positions (i.e., residue positions other than those included herein). Residue differences at these other residue positions can provide for additional variations in the amino acid sequence without adversely affecting the ability of the polypeptide to carry out the conversion of substrate to product. Accordingly, in some embodiments, in addition to the amino acid residue differences present in any one of the engineered TnTs polypeptides selected from the even-numbered sequences in the range of SEQ ID NO: 4-8 and 12, the sequence can further comprise 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 1-14, 1-15, 1-16, 1-18, 1-20, 1-22, 1-24, 1-26, 1-30, 1-35, 1-40, 1-45, 1-50, 1-100, or 1-150 residue differences at other amino acid residue positions as compared to the SEQ ID NOs: 4, 6, 8, or 12. In some embodiments, the number of amino acid residue differences as compared to the reference sequence can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 30, 35, 40, 45, 50, 100, or 150 residue positions. In some embodiments, the number of amino acid residue differences as compared to the reference sequence can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 20, 21, 22, 23, 24, or 25 residue positions. The residue differences at these other positions can be conservative changes or non-conservative changes. In some embodiments, the residue differences can comprise conservative substitutions and non-conservative substitutions as compared to the TnT polypeptide of SEQ ID NOs: 4, 6, 8, and/or 12.

[0179] In some embodiments, the present disclosure also provides engineered polypeptides that comprise a fragment of any of the engineered TnT polypeptides described herein that retains the functional activity and/or improved property of that engineered TnT. Accordingly, in some

embodiments, the present disclosure provides a polypeptide fragment capable of converting substrate to product under suitable reaction conditions, wherein the fragment comprises at least about 90%, 95%, 96%, 97%, 98%, or 99% of a full-length or truncated amino acid sequence of an engineered TnT of the present disclosure, such as an exemplary TnT polypeptide selected from the even-numbered sequences in the range of SEQ ID NO: 4-8 and 12. In some embodiments, the engineered TnT can have an amino acid sequence comprising a deletion in any one of the TnT polypeptide sequences described herein, such as the exemplary engineered polypeptides of the even-numbered sequences in the range of SEQ ID NO: 4-8 and 12.

[0180] Thus, for each and every embodiment of the engineered TnT polypeptides of the invention, the amino acid sequence can comprise deletions of one or more amino acids, 2 or more amino acids, 3 or more amino acids, 4 or more amino acids, 5 or more amino acids, 6 or more amino acids, 8 or more amino acids, 10 or more amino acids, 15 or more amino acids, or 20 or more amino acids, up to 10% of the total number of amino acids, up to 20% of the total number of amino acids, or up to 30% of the total number of amino acids of the TnT polypeptides, where the associated functional activity and/or improved properties of the engineered TnT described herein are maintained. In some embodiments, the deletions can comprise 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-15, 1-20, 1-21, 1-22, 1-23, 1-24, 1-25, 1-30, 1-35, 1-40, 1-45, or 1-50 amino acid residues. In some embodiments, the number of deletions can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 30, 35, 40, 45, or 50 amino acid residues. In some embodiments, the deletions can comprise deletions of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 20, 21, 22, 23, 24, or 25 amino acid residues.

[0181] In some embodiments, the engineered TnT polypeptide herein can have an amino acid sequence comprising an insertion as compared to any one of the engineered TnT polypeptides described herein, such as the exemplary engineered polypeptides of the even-numbered sequences in the range of SEQ ID NO: 4-8 and 12. Thus, for each and every embodiment of the TnT polypeptides of the invention, the insertions can comprise one or more amino acids, 2 or more amino acids, 3 or more amino acids, 4 or more amino acids, 5 or more amino acids, 6 or more amino acids, 8 or more amino acids, 10 or more amino acids, 15 or more amino acids, 20 or more amino acids, 30 or more amino acids, 40 or more amino acids, or 50 or more amino acids, where the associated functional activity and/or improved properties of the engineered TnT described herein is maintained. The insertions can be to amino or carboxy terminus, or internal portions of the TnT polypeptide.

[0182] In some embodiments, the engineered TnT described herein can have an amino acid sequence comprising a sequence selected from the even-numbered sequences in the range of SEQ ID NO: 4-8 and 12, and optionally one or several (e.g., up to 3, 4, 5, or up to 10) amino acid residue deletions, insertions and/or substitutions. In some embodiments, the amino acid sequence has optionally 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-15, 1-20, 1-21, 1-22, 1-23, 1-24, 1-25, 1-30, 1-35, 1-40, 1-45, 1-50, 1-75, 1-100, or 1-150 amino acid residue deletions, insertions and/or substitutions. In some

embodiments, the amino acid sequence has optionally around 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 30, 35, 40, 45, 50, 55, 60, 65, 70, 80, 90, 100, 110, 120, 130, 140, or 150 amino acid residue deletions, insertions and/or substitutions. In some embodiments, the substitutions can be conservative or non-conservative substitutions.

[0183] In some embodiments, the polypeptides of the present disclosure are fusion polypeptides in which the engineered polypeptides are fused to other polypeptides, such as, by way of example and not limitation, antibody tags (e.g., myc epitope), purification sequences (e.g., His tags for binding to metals), cell localization signals (e.g., secretion signals), and polypeptides with enzymatic activity. Thus, the engineered polypeptides described herein can be used with or without fusions to other polypeptides.

[0184] In one embodiment of the engineered TnT polypeptides of the present disclosure, the polypeptide further comprises an N-terminal truncation of 1-156 amino acids of the polypeptide sequence relative to any even-numbered sequence set forth in SEQ ID NO: 4-8 and 12.

[0185] In some embodiments, the engineered TnT polypeptides of the invention can be fused to another second polypeptide, such as a polypeptide with a different enzymatic activity. In some embodiments, the present provides a fusion polypeptide comprising an engineered TnT polypeptide fused to a second polypeptide with inorganic pyrophosphatase (IPP) activity. For example, synthetic genes encoding an *N*-terminal and *C*-terminal hexahistidine tagged version of a wild-type (WT) inorganic pyrophosphatase (IPP) polypeptide can be fused to gene encoding a TnT variant polypeptide. Typically, the polypeptides (e.g., IPP and TnT) are fused via a polypeptide linker (e.g., a GSGGTG linker) introduced in the construct between the genes encoding the polypeptides. Such fusion proteins can be constructed using well-established techniques (e.g., Gibson assembly cloning) and expressed in *E. coli* (e.g., a strain derived from W3110).

[0186] It is to be understood that the polypeptides described herein are not restricted to the genetically encoded amino acids. In addition to the genetically encoded amino acids, the polypeptides described herein may be comprised, either in whole or in part, of naturally occurring and/or synthetic non-encoded amino acids. Certain commonly encountered non-encoded amino acids of which the polypeptides described herein may be comprised include, but are not limited to: the D-stereoisomers of the genetically-encoded amino acids; 2,3-diaminopropionic acid (Dpr); α-aminoisobutyric acid (Aib); ε-aminohexanoic acid (Aha); δ-aminovaleric acid (Ava); N-methylglycine or sarcosine (MeGly or Sar); ornithine (Orn); citrulline (Cit); t-butylalanine (Bua); t-butylglycine (Bug); N-methylisoleucine (MeIle); phenylglycine (Phg); cyclohexylalanine (Cha); norleucine (Nle); naphthylalanine (Nal); 2-chlorophenylalanine (Ocf); 3-chlorophenylalanine (Mcf); 4-chlorophenylalanine (Pcf); 2-fluorophenylalanine (Off); 3-fluorophenylalanine (Mff); 4-fluorophenylalanine (Pff); 2-bromophenylalanine (Obf); 3-bromophenylalanine (Mbf); 4-

bromophenylalanine (Pbf); 2-methylphenylalanine (Omf); 3-methylphenylalanine (Mmf); 4-methylphenylalanine (Pmf); 2-nitrophenylalanine (Onf); 3-nitrophenylalanine (Mnf); 4-nitrophenylalanine (Pnf); 2-cyanophenylalanine (Ocf); 3-cyanophenylalanine (Mcf); 4-cyanophenylalanine (Pcf); 2-trifluoromethylphenylalanine (Otf); 3-trifluoromethylphenylalanine (Mtf); 4-trifluoromethylphenylalanine (Ptf); 4-aminophenylalanine (Paf); 4-iodophenylalanine (Pif); 4-aminomethylphenylalanine (Pamf); 2,4-dichlorophenylalanine (Opef); 3,4-dichlorophenylalanine (Mpcf); 2,4-difluorophenylalanine (Opff); 3,4-difluorophenylalanine (Mpff); pyrid-2-ylalanine (2pAla); pyrid-3-ylalanine (3pAla); pyrid-4-ylalanine (4pAla); naphth-1-ylalanine (1nAla); naphth-2-ylalanine (2nAla); thiazolylalanine (taAla); benzothienylalanine (bAla); thienylalanine (tAla); furylalanine (fAla); homophenylalanine (hPhe); homotyrosine (hTyr); homotryptophan (hTrp); pentafluorophenylalanine (5ff); styrylkalanine (sAla); authrylalanine (aAla); 3,3-diphenylalanine (Dfa); 3-amino-5-phenypentanoic acid (Afp); penicillamine (Pen); 1,2,3,4-tetrahydroisoquinoline-3-carboxylic acid (Tic); β-2-thienylalanine (Thi); methionine sulfoxide (Mso); N(w)-nitroarginine (nArg); homolysine (hLys); phosphonomethylphenylalanine (pmPhe); phosphoserine (pSer); phosphothreonine (pThr); homoaspartic acid (hAsp); homoglutanic acid (hGlu); 1-aminocyclopent-(2 or 3)-ene-4 carboxylic acid; pipecolic acid (PA), azetidine-3-carboxylic acid (ACA); 1-aminocyclopentane-3-carboxylic acid; allylglycine (aGly); propargylglycine (pgGly); homoalanine (hAla); norvaline (nVal); homoleucine (hLeu), homovaline (hVal); homoisoleucine (hIle); homoarginine (hArg); N-acetyl lysine (AcLys); 2,4-diaminobutyric acid (Dbu); 2,3-diaminobutyric acid (Dab); N-methylvaline (MeVal); homocysteine (hCys); homoserine (hSer); hydroxyproline (Hyp) and homoproline (hPro). Additional non-encoded amino acids of which the polypeptides described herein may be comprised will be apparent to those of skill in the art (See e.g., the various amino acids provided in Fasman, CRC Practical Handbook of Biochemistry and Molecular Biology, CRC Press, Boca Raton, FL, pp. 3-70 [1989], and the references cited therein, all of which are incorporated by reference). These amino acids may be in either the L- or D-configuration.

[0187] Those of skill in the art will recognize that amino acids or residues bearing side chain protecting groups may also comprise the polypeptides described herein. Non-limiting examples of such protected amino acids, which in this case belong to the aromatic category, include (protecting groups listed in parentheses), but are not limited to: Arg(tos), Cys(methylbenzyl), Cys (nitropyridinesulfenyl), Glu(δ-benzylester), Gln(xanthyl), Asn(N-δ-xanthyl), His(bom), His(benzyl), His(tos), Lys(fmoc), Lys(tos), Ser(O-benzyl), Thr (O-benzyl) and Tyr(O-benzyl).

[0188] Non-encoding amino acids that are conformationally constrained of which the polypeptides described herein may be composed include, but are not limited to, N-methyl amino acids (L-configuration); 1-aminocyclopent-(2 or 3)-ene-4-carboxylic acid; pipecolic acid; azetidine-3-carboxylic acid; homoproline (hPro); and 1-aminocyclopentane-3-carboxylic acid.

[0189] In some embodiments, the engineered polypeptides can be in various forms, for example, such as an isolated preparation, as a substantially purified enzyme, whole cells transformed with gene(s) encoding the enzyme, and/or as cell extracts and/or lysates of such cells. The enzymes can be lyophilized, spray-dried, precipitated or be in the form of a crude paste, as further discussed below.

[0190] In some embodiments, the engineered polypeptides can be in the form of a biocatalytic composition. In some embodiments, the biocatalytic composition comprises (a) a means for conversion of an NTP-3-O-RBG or natural or modified NTP comprising a conjugate moiety, reactive group, or linker substrate and an oligo acceptor compound comprising a conjugate moiety, reactive group, or linker to an oligo acceptor product extended by one nucleotide by contact with a TnT and (b) a suitable cofactor. The suitable cofactor may be cobalt, manganese, or any other suitable cofactor.

[0191] In some embodiments, the polypeptides described herein are provided in the form of kits. The enzymes in the kits may be present individually or as a plurality of enzymes. The kits can further include reagents for carrying out the enzymatic reactions, substrates for assessing the activity of enzymes, as well as reagents for detecting the products. The kits can also include reagent dispensers and instructions for use of the kits.

[0192] In some embodiments, the kits of the present disclosure include arrays comprising a plurality of different TnT polypeptides at different addressable position, wherein the different polypeptides are different variants of a reference sequence each having at least one different improved enzyme property. In some embodiments, a plurality of polypeptides immobilized on solid supports are configured on an array at various locations, addressable for robotic delivery of reagents, or by detection methods and/or instruments. The array can be used to test a variety of substrate compounds for conversion by the polypeptides. Such arrays comprising a plurality of engineered polypeptides and methods of their use are known in the art (See e.g., WO2009/008908A2).

Polynucleotides Encoding Engineered Terminal Nucleotidyl Transferases, Expression Vectors and Host Cells

[0193] In another aspect, the present disclosure provides polynucleotides encoding the engineered TnT polypeptides described herein. The polynucleotides may be operatively linked to one or more heterologous regulatory sequences that control gene expression to create a recombinant polynucleotide capable of expressing the polypeptide. Expression constructs containing a heterologous polynucleotide encoding the engineered TnT are introduced into appropriate host cells to express the corresponding TnT polypeptide.

[0194] As will be apparent to the skilled artisan, availability of a protein sequence and the knowledge of the codons corresponding to the various amino acids provide a description of all the polynucleotides capable of encoding the subject polypeptides. The degeneracy of the genetic code, where the same amino acids are encoded by alternative or synonymous codons, allows an extremely

large number of nucleic acids to be made, all of which encode the improved TnT enzymes. Thus, having knowledge of a particular amino acid sequence, those skilled in the art could make any number of different nucleic acids by simply modifying the sequence of one or more codons in a way which does not change the amino acid sequence of the protein. In this regard, the present disclosure specifically contemplates each and every possible variation of polynucleotides that could be made encoding the polypeptides described herein by selecting combinations based on the possible codon choices, and all such variations are to be considered specifically disclosed for any polypeptide described herein, including the amino acid sequences disclosed in the sequence listing incorporated by reference herein as the even-numbered sequences in the range of SEQ ID NO: 4-8 and 12.

[0195] In various embodiments, the codons are preferably selected to fit the host cell in which the protein is being produced. For example, preferred codons used in bacteria are used to express the gene in bacteria; preferred codons used in yeast are used for expression in yeast; and preferred codons used in mammals are used for expression in mammalian cells. In some embodiments, all codons need not be replaced to optimize the codon usage of the TnT since the natural sequence will comprise preferred codons and because use of preferred codons may not be required for all amino acid residues. Consequently, codon optimized polynucleotides encoding the TnT enzymes may contain preferred codons at about 40%, 50%, 60%, 70%, 80%, or greater than 90% of codon positions of the full-length coding region.

[0196] In some embodiments, the polynucleotide comprises a codon optimized nucleotide sequence encoding the TnT polypeptide amino acid sequence, as represented by SEQ ID NOs: 4, 6, 8, and 12. In some embodiments, the polynucleotide has a nucleic acid sequence comprising at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identity to the codon optimized nucleic acid sequences encoding the even-numbered sequences in the range of SEQ ID NOs: 4-8 and 12. In some embodiments, the polynucleotide has a nucleic acid sequence comprising at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identity to the codon optimized nucleic acid sequences in the odd-numbered sequences in the range of SEQ ID NOs: 3-7 and 11. In some embodiments, the codon optimized sequences of the odd-numbered sequences in the range of SEQ ID NOs: 3-7 and 11, enhance expression of the encoded TnT, providing preparations of enzyme capable of converting substrate to product.

[0197] In some embodiments, the polynucleotide sequence comprises at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to the reference sequence of SEQ ID NOs: 3, 5, 7, and/or 11 and/or or a functional fragment thereof, wherein said polynucleotide sequence encodes an engineered polypeptide comprising at least one substitution at one or more amino acid positions.

[0198] In some embodiments, the polynucleotide sequence encodes at least one engineered terminal nucleotidyl transferase comprising a sequence having at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to the reference sequence of SEQ ID NOs: 4, 6, 8, and 12.

[0199] In some embodiments, the polynucleotide sequence comprises SEQ ID NOs: 3, 5, 7, or 11.

[0200] In some embodiments, the polynucleotides are capable of hybridizing under highly stringent conditions to a reference sequence selected from the odd-numbered sequences in SEQ ID NOs: 3-7 and 11, or a complement thereof, and encode a TnT.

[0201] In some embodiments, as described above, the polynucleotide encodes an engineered TnT polypeptide with improved properties as compared to SEQ ID NOs: 4, 6, 8, 10, and/or 12, wherein the polypeptide comprises an amino acid sequence having at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identity to a reference sequence selected from SEQ ID NOs: 4, 6, 8, 10, and/or 12, and one or more residue differences as compared to SEQ ID NOs: 4, 6, 8, and/or 12, wherein the sequence is selected from the even-numbered sequences in the range of SEQ ID NOs: 4-8 and 12. In some embodiments, the reference amino acid sequence is selected from the even-numbered sequences in the range of SEQ ID NOs: 4-12. In some embodiments, the reference amino acid sequence is SEQ ID NO: 4; while in some other embodiments, the reference sequence is SEQ ID NO: 6; while in some other embodiments, the reference sequence is SEQ ID NO: 8. In some embodiments, the reference amino acid sequence is SEQ ID NO: 10; while in some other embodiments, the reference amino acid sequence is SEQ ID NO: 12.

[0202] In some embodiments, the polynucleotide encodes a TnT polypeptide capable of converting one or more substrates to product with improved properties as compared to SEQ ID NOs: 4, 6, 8, 10 and/or 12, wherein the polypeptide comprises an amino acid sequence having at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NOs: 4, 6, 8, 10, and/or 12.

[0203] In some embodiments, the polynucleotide encoding the engineered TnT comprises a polynucleotide sequence selected from the odd-numbered sequences in the range of SEQ ID NOs: 3-7 and 11.

[0204] In some embodiments, the polynucleotides are capable of hybridizing under highly stringent conditions to a reference polynucleotide sequence selected from the odd-numbered sequences in the range of SEQ ID NOs: 3-7 and 11 or a complement thereof and encode a TnT polypeptide with one or more of the improved properties described herein. In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes a TnT comprising an amino acid sequence having at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%,

99% or more identity to SEQ ID NOs: 4, 6, 8, 10, and/or 12, that has an amino acid sequence comprising one or more residue differences as compared to SEQ ID NOs: 4, 6, 8, 10, and/or 12, as described above and in the Examples, below.

[0205] In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes an engineered TnT polypeptide with improved properties comprising an amino acid sequence having at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identity to SEQ ID NOs: 4, 6, 8, 10, and/or 12. In some embodiments, the polynucleotides encode the polypeptides described herein but have at least about 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% or more sequence identity at the nucleotide level to a reference polynucleotide encoding the engineered TnT. In some embodiments, the reference polynucleotide sequence is selected from SEQ ID NOs: 3-11.

[0206] In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes an engineered TnT polypeptide with improved properties comprising an amino acid sequence having at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identity to SEQ ID NOs: 4, 6, 8, 10, and/or 12. In some embodiments, the polynucleotides encode the polypeptides described herein but have at least about 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% or more sequence identity at the nucleotide level to a reference polynucleotide encoding the engineered TnT. In some embodiments, the reference polynucleotide sequence is selected from SEQ ID NOs: 3-11.

[0207] In some embodiments, an isolated polynucleotide encoding any of the engineered TnT polypeptides provided herein is manipulated in a variety of ways to provide for expression of the polypeptide. In some embodiments, the polynucleotides encoding the polypeptides are provided as expression vectors where one or more control sequences is present to regulate the expression of the polynucleotides and/or polypeptides. Manipulation of the isolated polynucleotide prior to its insertion into a vector may be desirable or necessary depending on the expression vector. The techniques for modifying polynucleotides and nucleic acid sequences utilizing recombinant DNA methods are well known in the art.

[0208] In some embodiments, the control sequences include among other sequences, promoters, leader sequences, polyadenylation sequences, propeptide sequences, signal peptide sequences, and transcription terminators. As known in the art, suitable promoters can be selected based on the host cells used. For bacterial host cells, suitable promoters for directing transcription of the nucleic acid constructs of the present application, include, but are not limited to the promoters obtained from the *E. coli* lac operon, *Streptomyces coelicolor* agarase gene (dagA), *Bacillus subtilis* levansucrase gene (sacB), *Bacillus licheniformis* alpha-amylase gene (amyL), *Bacillus stearothermophilus* maltogenic amylase gene (amyM), *Bacillus amyloliquefaciens* alpha-amylase gene (amyQ), *Bacillus*

licheniformis penicillinase gene (penP), *Bacillus subtilis* xylA and xylB genes, and prokaryotic beta-lactamase gene (See e.g., Villa-Kamaroff et al., Proc. Natl Acad. Sci. USA 75: 3727-3731 [1978]), as well as the *tac* promoter (See e.g., DeBoer et al., Proc. Natl Acad. Sci. USA 80: 21-25 [1983]). Exemplary promoters for filamentous fungal host cells, include promoters obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Rhizomucor miehei* aspartic proteinase, *Aspergillus niger* neutral alpha-amylase, *Aspergillus niger* acid stable alpha-amylase, *Aspergillus niger* or *Aspergillus awamori* glucoamylase (glaA), *Rhizomucor miehei* lipase, *Aspergillus oryzae* alkaline protease, *Aspergillus oryzae* triose phosphate isomerase, *Aspergillus nidulans* acetamidase, and *Fusarium oxysporum* trypsin-like protease (See e.g., WO 96/00787), as well as the NA2-tpi promoter (a hybrid of the promoters from the genes for *Aspergillus niger* neutral alpha-amylase and *Aspergillus oryzae* triose phosphate isomerase), and mutant, truncated, and hybrid promoters thereof. Exemplary yeast cell promoters can be from the genes can be from the genes for *Saccharomyces cerevisiae* enolase (ENO-1), *Saccharomyces cerevisiae* galactokinase (GAL1), *Saccharomyces cerevisiae* alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase (ADH2/GAP), and *Saccharomyces cerevisiae* 3-phosphoglycerate kinase. Other useful promoters for yeast host cells are known in the art (See e.g., Romanos et al., Yeast 8:423-488 [1992]).

[0209] In some embodiments, the control sequence is a suitable transcription terminator sequence, a sequence recognized by a host cell to terminate transcription. The terminator sequence is operably linked to the 3' terminus of the nucleic acid sequence encoding the polypeptide. Any terminator which is functional in the host cell of choice finds use in the present disclosure. For example, exemplary transcription terminators for filamentous fungal host cells can be obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* glucoamylase, *Aspergillus nidulans* anthranilate synthase, *Aspergillus niger* alpha-glucosidase, and *Fusarium oxysporum* trypsin-like protease. Exemplary terminators for yeast host cells can be obtained from the genes for *Saccharomyces cerevisiae* enolase, *Saccharomyces cerevisiae* cytochrome C (CYC1), and *Saccharomyces cerevisiae* glyceraldehyde-3-phosphate dehydrogenase. Other useful terminators for yeast host cells are known in the art (See e.g., Romanos et al., *supra*).

[0210] In some embodiments, the control sequence is a suitable leader sequence, a non-translated region of an mRNA that is important for translation by the host cell. The leader sequence is operably linked to the 5' terminus of the nucleic acid sequence encoding the polypeptide. Any leader sequence that is functional in the host cell of choice may be used. Exemplary leaders for filamentous fungal host cells are obtained from the genes for *Aspergillus oryzae* TAKA amylase and *Aspergillus nidulans* triose phosphate isomerase. Suitable leaders for yeast host cells include but are not limited to those obtained from the genes for *Saccharomyces cerevisiae* enolase (ENO-1), *Saccharomyces cerevisiae* 3-phosphoglycerate kinase, *Saccharomyces cerevisiae* alpha-factor, and *Saccharomyces cerevisiae* alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase (ADH2/GAP). The control

sequence may also be a polyadenylation sequence, a sequence operably linked to the 3' terminus of the nucleic acid sequence and which, when transcribed, is recognized by the host cell as a signal to add polyadenosine residues to transcribed mRNA. Any polyadenylation sequence which is functional in the host cell of choice may be used in the present disclosure. Exemplary polyadenylation sequences for filamentous fungal host cells include but are not limited to those from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* glucoamylase, *Aspergillus nidulans* anthranilate synthase, *Fusarium oxysporum* trypsin-like protease, and *Aspergillus niger* alpha-glucosidase. Useful polyadenylation sequences for yeast host cells are also known in the art (See e.g., Guo and Sherman, Mol. Cell. Bio., 15:5983-5990 [1995]).

[0211] In some embodiments, the control sequence is a signal peptide coding region that codes for an amino acid sequence linked to the amino terminus of a polypeptide and directs the encoded polypeptide into the cell's secretory pathway. The 5' end of the coding sequence of the nucleic acid sequence may inherently contain a signal peptide coding region naturally linked in translation reading frame with the segment of the coding region that encodes the secreted polypeptide. Alternatively, the 5' end of the coding sequence may contain a signal peptide coding region that is foreign to the coding sequence. Any signal peptide coding region that directs the expressed polypeptide into the secretory pathway of a host cell of choice finds use for expression of the engineered TnT polypeptides provided herein. Effective signal peptide coding regions for bacterial host cells include but are not limited to the signal peptide coding regions obtained from the genes for *Bacillus* NClB 11837 maltogenic amylase, *Bacillus stearothermophilus* alpha-amylase, *Bacillus licheniformis* subtilisin, *Bacillus licheniformis* beta-lactamase, *Bacillus stearothermophilus* neutral proteases (nprT, nprS, nprM), and *Bacillus subtilis* prsA. Further signal peptides are known in the art (See e.g., Simonen and Palva, Microbiol. Rev., 57:109-137 [1993]). Effective signal peptide coding regions for filamentous fungal host cells include but are not limited to the signal peptide coding regions obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* neutral amylase, *Aspergillus niger* glucoamylase, *Rhizomucor miehei* aspartic proteinase, *Humicola insolens* cellulase, and *Humicola lanuginosa* lipase. Useful signal peptides for yeast host cells include but are not limited to those from the genes for *Saccharomyces cerevisiae* alpha-factor and *Saccharomyces cerevisiae* invertase.

[0212] In some embodiments, the control sequence is a propeptide coding region that codes for an amino acid sequence positioned at the amino terminus of a polypeptide. The resultant polypeptide is referred to as a "proenzyme," "propolypeptide," or "zymogen," in some cases). A propolypeptide can be converted to a mature active polypeptide by catalytic or autocatalytic cleavage of the propeptide from the propolypeptide. The propeptide coding region includes but is not limited to the genes for *Bacillus subtilis* alkaline protease (aprE), *Bacillus subtilis* neutral protease (nprT), *Saccharomyces cerevisiae* alpha-factor, *Rhizomucor miehei* aspartic proteinase, and *Myceliophthora thermophila* lactase (See e.g., WO 95/33836). Where both signal peptide and propeptide regions are present at the

amino terminus of a polypeptide, the propeptide region is positioned next to the amino terminus of a polypeptide and the signal peptide region is positioned next to the amino terminus of the propeptide region.

[0213] In some embodiments, regulatory sequences are also utilized. These sequences facilitate the regulation of the expression of the polypeptide relative to the growth of the host cell. Examples of regulatory systems are those which cause the expression of the gene to be turned on or off in response to a chemical or physical stimulus, including the presence of a regulatory compound. In prokaryotic host cells, suitable regulatory sequences include, but are not limited to the *lac*, *tac*, and *trp* operator systems. In yeast host cells, suitable regulatory systems include, but are not limited to the ADH2 system or GAL1 system. In filamentous fungi, suitable regulatory sequences include, but are not limited to the TAKA alpha-amylase promoter, *Aspergillus niger* glucoamylase promoter, and *Aspergillus oryzae* glucoamylase promoter.

[0214] The present disclosure also provides recombinant expression vectors comprising a polynucleotide encoding an engineered TnT polypeptide, and one or more expression regulating regions such as a promoter and a terminator, a replication origin, etc., depending on the type of hosts into which they are to be introduced. In some embodiments, the various nucleic acid and control sequences described above are combined together to produce a recombinant expression vector which includes one or more convenient restriction sites to allow for insertion or substitution of the nucleic acid sequence encoding the variant TnT polypeptide at such sites. Alternatively, the polynucleotide sequence(s) of the present disclosure are expressed by inserting the polynucleotide sequence or a nucleic acid construct comprising the polynucleotide sequence into an appropriate vector for expression. In creating the expression vector, the coding sequence is located in the vector so that the coding sequence is operably linked with the appropriate control sequences for expression.

[0215] The recombinant expression vector may be any vector (e.g., a plasmid or virus), that can be conveniently subjected to recombinant DNA procedures and can result in the expression of the variant TnT polynucleotide sequence. The choice of the vector will typically depend on the compatibility of the vector with the host cell into which the vector is to be introduced. The vectors may be linear or closed circular plasmids.

[0216] In some embodiments, the expression vector is an autonomously replicating vector (i.e., a vector that exists as an extra-chromosomal entity, the replication of which is independent of chromosomal replication, such as a plasmid, an extra-chromosomal element, a minichromosome, or an artificial chromosome). The vector may contain any means for assuring self-replication. In some alternative embodiments, the vector may be one which, when introduced into the host cell, is integrated into the genome and replicated together with the chromosome(s) into which it has been

integrated. Furthermore, a single vector or plasmid or two or more vectors or plasmids which together contain the total DNA to be introduced into the genome of the host cell, or a transposon may be used.

[0217] In some embodiments, the expression vector preferably contains one or more selectable markers, which permit easy selection of transformed cells. A "selectable marker" is a gene the product of which provides for biocide or viral resistance, resistance to heavy metals, prototrophy to auxotrophy, and the like. Examples of bacterial selectable markers include but are not limited to the *dal* genes from *Bacillus subtilis* or *Bacillus licheniformis*, or markers, which confer antibiotic resistance such as ampicillin, kanamycin, chloramphenicol or tetracycline resistance. Suitable markers for yeast host cells include, but are not limited to ADE2, HIS3, LEU2, LYS2, MET3, TRP1, and URA3. Selectable markers for use in a filamentous fungal host cell include, but are not limited to, amdS (acetamidase), argB (ornithine carbamoyltransferases), bar (phosphinothricin acetyltransferase), hph (hygromycin phosphotransferase), niaD (nitrate reductase), pyrG (orotidine-5'-phosphate decarboxylase), sC (sulfate adenyltransferase), and trpC (anthranilate synthase), as well as equivalents thereof. In another aspect, the present disclosure provides a host cell comprising a polynucleotide encoding at least one engineered TnT polypeptide of the present disclosure, the polynucleotide being operatively linked to one or more control sequences for expression of the engineered TnT enzyme(s) in the host cell. Host cells for use in expressing the polypeptides encoded by the expression vectors of the present disclosure are well known in the art and include but are not limited to, bacterial cells, such as *E. coli*, *Vibrio fluvialis*, *Streptomyces* and *Salmonella typhimurium* cells; fungal cells, such as yeast cells (e.g., *Saccharomyces cerevisiae* and *Pichia pastoris* [ATCC Accession No. 201178]); insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, BHK, 293, and Bowes melanoma cells; and plant cells. Exemplary host cells are *Escherichia coli* strains (e.g., W3110 (ΔfhuA) and BL21).

[0218] In some embodiments, the host cell strain comprises a knockout of one or more genes, in particular phosphatase genes. In some embodiments, the host cell comprises a knockout or single gene deletion of *E. coli genes aphA*, *surE*, *phoA*, and/or *cpdB*, as described below in the Examples. In some embodiments, the host cell comprising a knockout of one or more phosphatase genes has increased production of the product and/or decreased de-phosphorylation of the product or substrate.

[0219] Accordingly, in another aspect, the present disclosure provides methods for producing the engineered TnT polypeptides, where the methods comprise culturing a host cell capable of expressing a polynucleotide encoding the engineered TnT polypeptide under conditions suitable for expression of the polypeptide. In some embodiments, the methods further comprise the steps of isolating and/or purifying the TnT polypeptides, as described herein.

[0220] Appropriate culture media and growth conditions for the above-described host cells are well known in the art. Polynucleotides for expression of the TnT polypeptides may be introduced into cells

by various methods known in the art. Techniques include, among others, electroporation, biolistic particle bombardment, liposome mediated transfection, calcium chloride transfection, and protoplast fusion.

[0221] The engineered TnTs with the properties disclosed herein can be obtained by subjecting the polynucleotide encoding the naturally occurring or engineered TnT polypeptide to mutagenesis and/or directed evolution methods known in the art, and as described herein. An exemplary directed evolution technique is mutagenesis and/or DNA shuffling (See e.g., Stemmer, Proc. Natl. Acad. Sci. USA 91:10747-10751 [1994]; WO 95/22625; WO 97/0078; WO 97/35966; WO 98/27230; WO 00/42651; WO 01/75767 and U.S. Pat. 6,537,746). Other directed evolution procedures that can be used include, among others, staggered extension process (StEP), in vitro recombination (See e.g., Zhao et al., Nat. Biotechnol., 16:258–261 [1998]), mutagenic PCR (See e.g., Caldwell et al., PCR Methods Appl., 3:S136-S140 [1994]), and cassette mutagenesis (See e.g., Black et al., Proc. Natl. Acad. Sci. USA 93:3525-3529 [1996]).

[0222] For example, mutagenesis and directed evolution methods can be readily applied to polynucleotides to generate variant libraries that can be expressed, screened, and assayed. Mutagenesis and directed evolution methods are well known in the art (See e.g., US Patent Nos. 5,605,793, 5,811,238, 5,830,721, 5,834,252, 5,837,458, 5,928,905, 6,096,548, 6,117,679, 6,132,970, 6,165,793, 6,180,406, 6,251,674, 6,265,201, 6,277,638, 6,287,861, 6,287,862, 6,291,242, 6,297,053, 6,303,344, 6,309,883, 6,319,713, 6,319,714, 6,323,030, 6,326,204, 6,335,160, 6,335,198, 6,344,356, 6,352,859, 6,355,484, 6,358,740, 6,358,742, 6,365,377, 6,365,408, 6,368,861, 6,372,497, 6,337,186, 6,376,246, 6,379,964, 6,387,702, 6,391,552, 6,391,640, 6,395,547, 6,406,855, 6,406,910, 6,413,745, 6,413,774, 6,420,175, 6,423,542, 6,426,224, 6,436,675, 6,444,468, 6,455,253, 6,479,652, 6,482,647, 6,483,011, 6,484,105, 6,489,146, 6,500,617, 6,500,639, 6,506,602, 6,506,603, 6,518,065, 6,519,065, 6,521,453, 6,528,311, 6,537,746, 6,573,098, 6,576,467, 6,579,678, 6,586,182, 6,602,986, 6,605,430, 6,613,514, 6,653,072, 6,686,515, 6,703,240, 6,716,631, 6,825,001, 6,902,922, 6,917,882, 6,946,296, 6,961,664, 6,995,017, 7,024,312, 7,058,515, 7,105,297, 7,148,054, 7,220,566, 7,288,375, 7,384,387, 7,421,347, 7,430,477, 7,462,469, 7,534,564, 7,620,500, 7,620,502, 7,629,170, 7,702,464, 7,747,391, 7,747,393, 7,751,986, 7,776,598, 7,783,428, 7,795,030, 7,853,410, 7,868,138, 7,783,428, 7,873,477, 7,873,499, 7,904,249, 7,957,912, 7,981,614, 8,014,961, 8,029,988, 8,048,674, 8,058,001, 8,076,138, 8,108,150, 8,170,806, 8,224,580, 8,377,681, 8,383,346, 8,457,903, 8,504,498, 8,589,085, 8,762,066, 8,768,871, 9,593,326, and all related US, as well as PCT and non-US counterparts; Ling et al., Anal. Biochem., 254(2):157-78 [1997]; Dale et al., Meth. Mol. Biol., 57:369-74 [1996]; Smith, Ann. Rev. Genet., 19:423-462 [1985]; Botstein et al., Science, 229:1193-1201 [1985]; Carter, Biochem. J., 237:1-7 [1986]; Kramer et al., Cell, 38:879-887 [1984]; Wells et al., Gene, 34:315-323 [1985]; Minshull et al., Curr. Op. Chem. Biol., 3:284-290 [1999]; Christians et al., Nat. Biotechnol., 17:259-264 [1999]; Crameri et al., Nature, 391:288-291 [1998]; Crameri, et al., Nat. Biotechnol., 15:436-438

[1997]; Zhang et al., Proc. Nat. Acad. Sci. U.S.A., 94:4504-4509 [1997]; Crameri et al., Nat. Biotechnol., 14:315-319 [1996]; Stemmer, Nature, 370:389-391 [1994]; Stemmer, Proc. Nat. Acad. Sci. USA, 91:10747-10751 [1994]; WO 95/22625; WO 97/0078; WO 97/35966; WO 98/27230; WO 00/42651; WO 01/75767; and WO 2009/152336, all of which are incorporated herein by reference).

[0223] In some embodiments, the enzyme clones obtained following mutagenesis treatment are screened by subjecting the enzymes to a defined temperature (or other assay conditions, such as testing the enzyme's activity over a broad range of substrates) and measuring the amount of enzyme activity remaining after heat treatments or other assay conditions. Clones containing a polynucleotide encoding a TnT polypeptide are then sequenced to identify the nucleotide sequence changes (if any) and used to express the enzyme in a host cell. Measuring enzyme activity from the expression libraries can be performed using any suitable method known in the art (e.g., standard biochemistry techniques, such as HPLC analysis).

[0224] In some embodiments, the clones obtained following mutagenesis treatment can be screened for engineered TnTs having one or more desired improved enzyme properties (e.g., improved regioselectivity). Measuring enzyme activity from the expression libraries can be performed using the standard biochemistry techniques, such as HPLC analysis, LC-MS analysis, RapidFire-MS analysis, and/or capillary electrophoresis analysis.

[0225] When the sequence of the engineered polypeptide is known, the polynucleotides encoding the enzyme can be prepared by standard solid-phase methods, according to known synthetic methods. In some embodiments, fragments of up to about 100 bases can be individually synthesized, then joined (*e.g.*, by enzymatic or chemical ligation methods, or polymerase mediated methods) to form any desired continuous sequence. For example, polynucleotides and oligonucleotides encoding portions of the TnT can be prepared by chemical synthesis as known in the art (e.g., the classical phosphoramidite method of Beaucage et al., Tet. Lett. 22:1859-69 [1981], or the method described by Matthes et al., EMBO J. 3:801-05 [1984]) as typically practiced in automated synthetic methods. According to the phosphoramidite method, oligonucleotides are synthesized (*e.g.*, in an automatic DNA synthesizer), purified, annealed, ligated and cloned in appropriate vectors. In addition, essentially any nucleic acid can be obtained from any of a variety of commercial sources. In some embodiments, additional variations can be created by synthesizing oligonucleotides containing deletions, insertions, and/or substitutions, and combining the oligonucleotides in various permutations to create engineered TnTs with improved properties.

[0226] Accordingly, in some embodiments, a method for preparing the engineered TnT polypeptide comprises: (a) synthesizing a polynucleotide encoding a polypeptide comprising an amino acid sequence having at least about 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% or more sequence identity to an amino acid sequence selected from the even-

numbered sequences of SEQ ID NOs: 4-12, and having one or more residue differences as compared to SEQ ID NOs: 4, 6, 8, 10 and/or 12; and (b) expressing the TnT polypeptide encoded by the polynucleotide.

[0227] In some embodiments of the method, the polynucleotide encodes an engineered TnT that has optionally one or several (e.g., up to 3, 4, 5, or up to 10) amino acid residue deletions, insertions and/or substitutions. In some embodiments, the amino acid sequence has optionally 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-15, 1-20, 1-21, 1-22, 1-23, 1-24, 1-25, 1-30, 1-35, 1-40, 1-45, 1-50, 1-75, 1-100, or 1-150 amino acid residue deletions, insertions and/or substitutions. In some embodiments, the amino acid sequence has optionally around 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 30, 35, 40, 45, 50, 55, 60, 65, 70, 80, 90, 100, 110, 120, 130, 140, or 150 amino acid residue deletions, insertions and/or substitutions. In some embodiments, the substitutions can be conservative or non-conservative substitutions.

[0228] In some embodiments, any of the engineered TnT enzymes expressed in a host cell can be recovered from the cells and/or the culture medium using any one or more of the well-known techniques for protein purification, including, among others, lysozyme treatment, sonication, filtration, salting-out, ultra-centrifugation, and chromatography. Suitable solutions for lysing and the high efficiency extraction of proteins from bacteria, such as *E. coli*, are commercially available (e.g., CelLytic B™, Sigma-Aldrich, St. Louis MO).

[0229] Chromatographic techniques for isolation of the TnT polypeptide include, among others, reverse phase chromatography high performance liquid chromatography, ion exchange chromatography, gel electrophoresis, and affinity chromatography. Conditions for purifying a particular enzyme will depend, in part, on factors such as net charge, hydrophobicity, hydrophilicity, molecular weight, molecular shape, etc., and will be apparent to those having skill in the art.

[0230] In some embodiments, affinity techniques may be used to isolate the improved TnT enzymes. For affinity chromatography purification, any antibody which specifically binds the TnT polypeptide may be used. For the production of antibodies, various host animals, including but not limited to rabbits, mice, rats, etc., may be immunized by injection with a TnT polypeptide, or a fragment thereof. The TnT polypeptide or fragment may be attached to a suitable carrier, such as BSA, by means of a side chain functional group or linkers attached to a side chain functional group. In some embodiments, the affinity purification can use a specific ligand bound by the TnT or dye affinity column (See e.g., EP0641862; Stellwagen, "Dye Affinity Chromatography," *In* Current Protocols in Protein Science, Unit 9.2-9.2.16 [2001]).

Methods of Using the Engineered TnT Enzymes

[0231] In some embodiments, the TnT enzymes described herein find use in processes for conversion of one or more suitable substrates to a product.

[0232] In some embodiments, the engineered TnT polypeptides disclosed herein can be used in a process for the conversion of the oligo acceptor substrate and an NTP-3'-O-RBG or natural or modified NTP substrate to a product comprising an oligo acceptor substrate extended by one nucleotide, wherein either the oligo acceptor substrate or the natural or modified NTP substrate or both comprise a conjugate moiety, reactive group, or linker.

[0233] In some embodiments, the TnT polypeptides disclosed herein have activity incorporating a NTP-3'-O-RBG or natural or modified NTP substrate comprising a conjugate moiety, reactive group, or linker onto a polynucleotide or oligo acceptor substrate and/or are capable of adding additional NTP-3'-O-RBGs or natural or modified NTP substrates onto a polynucleotide or oligo acceptor substrate comprising one or more conjugate moieties, reactive groups, or linkers.

[0234] In the embodiments provided herein and illustrated in the Examples, various ranges of suitable reaction conditions that can be used in the processes, include but are not limited to, substrate loading, co-substrate loading, pH, temperature, buffer, solvent system, cofactor, polypeptide loading, and reaction time. Further suitable reaction conditions for carrying out the process for biocatalytic conversion of substrate compounds to product compounds using an engineered TnT described herein can be readily optimized in view of the guidance provided herein by routine experimentation that includes, but is not limited to, contacting the engineered TnT polypeptide and one or more substrate compounds under experimental reaction conditions of concentration, pH, temperature, and solvent conditions, and detecting the product compound.

Oligo Acceptor Substrate

[0235] The oligo acceptor substrate may be any nucleotide chain or similar moiety with an exposed 3'-OH. In some embodiments, the acceptor substrate may be single stranded. In yet other embodiments, the acceptor substrate may be double stranded or partially doubled stranded. In yet some other embodiments, the oligo acceptor substrate comprises a hairpin or circular structure. In some embodiments, the acceptor substrate may comprise a nucleotide chain consisting of 1-10 nucleotides, 5-20 nucleotides, 15-50 nucleotides, 30-100 nucleotides, or greater than 100 nucleotides. In some embodiments, the oligo acceptor substrate may comprise a chemical moiety that is not a nucleotide chain but contains a free -OH capable of being recognized as a substrate by a wild-type or engineered TnT.

[0236] In some embodiments, the oligo acceptor substrate comprises a nucleotide chain of repeating nucleotides. In other embodiments, the oligo acceptor substrate comprises a nucleotide chain of varied nucleotides that do not repeat. In some embodiments, the oligo acceptor substrate comprises a nucleotide chain with an odd number of nucleotides. In some embodiments, the oligo acceptor substrate comprises a nucleotide with an even number of nucleotides.

[0237] In some embodiments, the oligo acceptor substate is secured to a solid support. Suitable solid supports are known to those in the art and described, below, in this disclosure.

[0238] In some embodiments, the oligo acceptor substrate comprises one or more nucleotide sequences selected from the following 5'-6-FAM-T10mCmCmUfA, 5'-6-FAM-T11AmC*mA*mG, 5'-6-FAM-T11mU*fA*fA, 5'-6-FAM-T15mAmUmCmU, 5'-6-FAM-T16mC*mA*mGmA, 5'-6-FAM-T16mGmUmC*mC, 5'-6-FAM-T17*fA*fAfG, 5'-6-FAM-T21mGfUfAfC, 5'-6-FAM-T26mCfCmCfG, 5'-6-FAM-T26mUfGmUfC, 5'-6-FAM-T27fGmAfU, 5'-6-FAM-T31mAmAmAfG, 5'-6-FAM-T31mUfCmAfU, 5'-6-FAM-T36fCmAfUfC, 5'-6-FAM-T36fCmAmUmC, 5'-6-FAM-T36mC*mU*mAmC, 5'-6-FAM-T36mCfCmGfG, 5'-6-FAM-T41*fC*mAfA, 5'-6-FAM-T41mAfUfCfC, 5'-6-FAM-T41mAmUmCmU, 5'-6-FAM-T41mGfUmGfG, 5'-6-FAM-T46fAmUmUfG, 5'-6-FAM-T46mAfGmUfG, 5'-6-FAM-T46mUmCmUmU, 5'-6-FAM-T51fGmUfGmU, 5'-6-FAM-T51mU*fA*fAfG, 5'-6-FAM-T56mA*fC*mAfA, 5'-6-FAM-T9mAmAmAmUmCmU, 5'P-mUmCmU, mAfGmUfG, mAmAmAmAmUmCmU, mAmAmAmUmCmU, mAmAmUmCmU, mAmUmCmU, mC*mA*mGmA, mGmUmC*mC, mUmCmU, T8fCmAmUmC, T8mCfCmCfG, T8mCfCmGfG, T8mUfGmUfC. These embodiments are intended to be non-limiting. Any suitable oligo acceptor substrate finds use in the present disclosure.

Nucleotide Triphosphate, Modified Nucleotide Triphosphate, or NTP-3'-O-RBG Substrate

[0239] In some embodiments, the nucleotide triphosphate, modified nucleotide triphosphate, or NTP-3'-O-RBG substrate comprises a deoxyribonucleoside triphosphate with a 3'-O-RBG. In other embodiments, the nucleotide triphosphate, modified nucleotide triphosphate, or NTP-3'-O-RBG substrate may comprise a ribonucleoside triphosphate with a 3'-O-RBG. In yet other embodiments, the nucleotide triphosphate, modified nucleotide triphosphate, or NTP-3'-O-RBG substrate may comprise a synthetic nucleoside triphosphate with a 3'-O-RBG. In some embodiments, the NTP-3'-O-RBG substrate may comprise a sugar ring with a number of carbons that is not five. A non-limiting example of this is a threose nucleoside triphosphate.

[0240] A range of 3' removable blocking groups for the nucleotide triphosphate, modified nucleotide triphosphate, or NTP-3'-O-RBG substrate useful in the present disclosure are known in the art and include but are not limited to, NH_2, $-NO_2$, $-(CH_2)_2-CN$, or $-PO_3$. In some embodiments, the 3' removable blocking group of the nucleotide triphosphate, modified nucleotide triphosphate, or NTP-3'-O-RBG substrate is a carbonitrile, phosphate, carbonate, carbamate, ester, ether, borate, nitrate, sugar, phosphoramidate, phenylsulfenate, or sulfate. In some embodiments, the NTP-3'-O-RBG substrate with a 3' removable blocking group can be selected from the group consisting of NTP-3'-O-NH_2, NTP-3'-O-NO_2, or NTP-3'-O-PO_3. In some embodiments, the NTP-3'-O-RBG substrate comprises another blocking group that would sterically hinder addition of a second NTP-3'-O-RBG

substrate to the 3' end of the growing oligo acceptor substrate strand prior to removal of the removable blocking from the first round of addition.

[0241] In some embodiments, the nucleotide triphosphate, modified nucleotide triphosphate, or NTP-3'-O-RBG substrate comprises one or more conjugate moieties, reactive groups, or linkers, as further described, below.

Further Substrate Modifications

[0242] In some further embodiments, either the oligo acceptor substrate or the nucleotide triphosphate, modified nucleotide triphosphate, or NTP-3'-O-RBG substrate or both may comprise one or more modifications. Modifications may be at the 3' position, as is in the case of NTP-3'-O-RBG, or at the 2' position. Modifications may be at other positions of the sugar or to the base. Modifications may also be present as substitutions of one or more of the phosphate groups of the nucleotide triphosphate and may be incorporated into the phospho backbone of the growing oligonucleotide change. Although specific examples of suitable modifications are provided herein, any modification to the nucleotide triphosphate may be used in the described methods. Various modifications may confer various desired properties to the oligonucleotide chain.

[0243] In some embodiments, either the oligo acceptor substrate or the nucleotide triphosphate, modified nucleotide triphosphate, or NTP-3'-O-RBG substrate or both may comprise one or more nucleotides with a 2' modification, as described herein. In some embodiments, either the oligo acceptor substrate or the nucleotide triphosphate, modified nucleotide triphosphate, or NTP-3'-O-RBG substrate or both may comprise one or more nucleotides with a 2' modification selected from 2'-OH, 2'-H, 2'-O-methyl, 2'-fluoro, or 2'-O-2-methoxyethyl, 2'-$OCH_2CH_2OCH_3$, 2'-CO_2R' (where R' is any alkyl or aryl), or another 2' atom or chemical group. In some embodiments, the 2' modification comprises a 2'-F or 2'-O-alkyl. In some further embodiments, the 2'-F modified nucleotide comprises 2'-fluoro-2'-deoxyadenosine-5'-triphosphate, 2'-fluoro-2'-deoxycytidine-5'-triphosphate, 2'-fluoro-2'-deoxyguanosine-5'-triphosphate, and 2'-fluoro-2'-deoxyuridine-5'-triphosphate. In some further embodiments, the 2'-O-alkyl modified nucleotide comprises 2'-O-methyladenosine-5'- triphosphate, 2'-O-methylcytidine-5'-triphosphate, 2'-O-methylguanosine-5'-triphosphate, 2'- O-methyluridine-5'-triphosphate, and 2'-O-methylinosine-5'-triphosphate.

[0244] In some embodiments, either the oligo acceptor substrate or the nucleotide triphosphate, modified nucleotide triphosphate, or NTP-3'-O-RBG substrate or both may comprise one or more additional modifications, such as a phosphorothioate. In some embodiments, the modified nucleotide triphosphate comprises or further comprises a phosphorothioate group at the 5' alpha position. In some embodiments, the modified nucleotide triphosphate further comprises substitution of two oxygens for sulfurs for the creation of phosphorodithioate linkages.

[0245] In some embodiments, the sugar may have other modifications at other positions, such as locked nucleotides or constrained ethyl nucleotides, as is known in the art. In some embodiments, "locked nucleoside" or "locked nucleotide" or "locked nucleic acid" (LNA) refers to a nucleoside, nucleotide or nucleic acid, respectively, in which the ribose moiety is modified with a bridge connecting the 2' oxygen and 4' carbon. The LNA may be either in the C3'-endo (beta-D-LNA) or C2'-endo (alpha-L-LNA) conformation (see, e.g., Obika et al., Tetrahedron Letters, 1997, 38(50):8735–8738; Orum et al., Current Pharmaceutical Design, 2008, 14(11):1138–1142). Typically, the bridge is a methylene bridge. LNA may confer additional stability to a polynucleotide.

[0246] In some embodiments, either the oligo acceptor substrate or the nucleotide triphosphate, modified nucleotide triphosphate, or NTP-3'-O-RBG substrate or both may comprise a glycol nucleic acid (GNA) modification, wherein the sugar moiety of the NTP is replaced by propylene glycol. GNA may confer additional stability to a polynucleotide.

[0247] In some embodiments, either the oligo acceptor substrate or the nucleotide triphosphate, modified nucleotide triphosphate, or NTP-3'-O-RBG substrate or both may comprise one or more natural purine or pyrimidine bases, such as adenine, guanine, cytosine, thymine, or uridine. In some embodiments, either the oligo acceptor substrate or the nucleotide triphosphate, modified nucleotide triphosphate, or NTP-3'-O-RBG substrate or both may comprise one or more unnatural base analogs such as inosine, xanthine, hypoxanthine, or another base analog, as is known in the art. Various modified nucleobases or base analogs are known to those skilled in the art, including but not limited to the following: 5-methylcytosine, pyridin-4-one, pyridin-2-one, phenyl, pseudouracil, 3-methyl uracil, dihydrouridine, naphthyl, aminophenyl, 5- alkylcytidines, 5-alkyluridines, 5-halouridines, 6-azapyrimidines, 6-alkylpyrimidines, propyne, quesosine, 2-thiouridine, 4-thiouridine, 4-acetyltidine, 5- (carboxyhydroxymethyl)uridine, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluridine, -D-galactosylqueosine, 1-methyladenosine, 1- methylinosine, 2,2-dimethylguanosine, 3-methylcytidine, 2-methyladenosine, 2- methylguanosine, N6-methyladenosine, 7-methylguanosine, 5-methoxyaminomethy1-2- thiouridine, 5-methylaminomethy luridine, 5-methylcarbon ylmethyluridine, 5- methyloxyuridine, 5-methyl-2-thiouridine, 2-methylthio-N6-isopentenyladenosine, -D¬ mannosylqueosine, uridine-5-oxyacetic acid, 2-thiocytidine, N1-methyl-adenine, N6-methyl¬ adenine, 8'-azido-adenine, N,N-dimethy1-adenosine, aminoally1-adenosine, 5'-methy1-urdine, pseudouridine, N1- methy1-pseudouridine, 5'-hydroxy-methy1-uridine, 2'-thio-uridine, 4'-thio¬ uridine, hypoxanthine, xanthine, 5'-methyl-cytidine, 5'-hydroxy-methyl-cytidine, 6'-thio¬ guanine, and N7-methyl-guanine.

Conjugate Moieties

[0248] In some embodiments, either the oligo acceptor substrate or the nucleotide triphosphate, modified nucleotide triphosphate, or NTP-3'-O-RBG substrate or both comprises a conjugate moiety.

[0249] In some embodiments, the conjugate moiety (i.e., non-nucleotide moiety) includes, among others, carbohydrates (e.g. GalNAc), lipids, sterols, drug substances, hormones, polymers (e.g., polyethylene glycol, etc.), proteins, peptides, toxins (e.g. bacterial toxins, etc.), vitamins (e.g., folate, tocopherol, retinoic acid, etc.), or combinations thereof. In some embodiments, the conjugate moiety is used to affect the pharmacokinetics of the oligonucleotide and/or oligonucleotide cell targeting.

[0250] In some embodiments, the conjugate moiety can be attached to the 5'-terminal nucleotide, the 3'-terminal nucleotide, or an internal nucleotide. In some embodiments, the conjugate moiety is attached to the 2'-position of the sugar moiety of a nucleoside, for example, to the 2'-OH. In some embodiments, the conjugate moiety is attached to the 3'-position of the sugar moiety of the nucleoside, for example 3'-OH. In some embodiments, the conjugate moiety is attached to the nucleobase, (see, e.g., Biscans et al., Nucleic Acids Res. 2019 Feb 20; 47(3): 1082–1096). In some embodiments, the conjugate moiety is attached directly or attached using a linker.

[0251] In some embodiments, the conjugate moiety comprises a C_6-C_{22} alkyl, C_{6-22} alkenyl, or C_6-C_{22} alkynyl. In some embodiments, the conjugate moiety comprises a C_6-alkyl, C_7-alkyl, C_8-alkyl, C_9-alkyl, C_{10}-alkyl, C_{11}-alkyl, C_{12}-alkyl, C_{13}-alkyl, C_{14}-alkyl, C_{15}-alkyl, C_{16}-alkyl, C_{17}-alkyl, C_{18}-alkyl, C_{19}-alkyl, C_{20}-alkyl, C_{21}-alkyl, or C_{22}-alkyl. In some embodiments, the conjugate moiety comprises a C_6 alkenyl, C_7 alkenyl, C_8 alkenyl C_9 alkenyl, C_{10} alkenyl, C_{11}-alkenyl, C_{12}-alkenyl, C_{13}-alkenyl, C_{14}-alkenyl, C_{15}-alkenyl, C_{16}-alkenyl, C_{17}-alkenyl, C_{18}-alkenyl, C_{19}-alkenyl, C_{20}-alkenyl, C_{21}-alkenyl, or C_{22}-alkenyl. In some embodiments, the conjugate moiety comprises a C_6 alkynyl, C_7 alkynyl, C_8 alkynyl, C_9 alkynyl, C_{10} alkynyl, C_{11}-alkynyl, C_{12}-alkynyl, C_{13}-alkynyl, C_{14}-alkynyl, C_{15}-alkynyl, C_{16}-alkynyl, C_{17}-alkynyl, C_{18}-alkynyl, C_{19}-alkynyl, C_{20}-alkynyl, C_{21}-alkynyl, or C_{22}-alkynyl.

[0252] In some embodiments, the conjugate moiety comprises a heteroalkyl, heteroalkenyl, or heteroalkynyl. In some embodiments, the heteroalkyl, heteroalkenyl or heteroalkynyl has one or more carbon atoms replaced with a heteroatom, such as O, S, or N.

[0253] In some embodiments, the conjugate moiety comprises a cycloalkyl or heterocycloalkyl group. In some embodiments, the cycloalkyl includes, among others, cyclopropyl, cyclobutyl, cyclopentyl, cyclohexyl, l-cyclohexenyl, 3-cyclohexenyl, and cycloheptyl. In some embodiments, the heterocycloalkyl includes, among others, 1-(1,2,5,6-tetrahydropyridyfh l-piperidinyl, 2-piperidinyl, 3-piperidinyl, 4-morpholinyl, 3-morpholinyl, tctrahydrofuran-2-yl, tctrahydrofuran-3-yl, tetrahydrothicn-2-yl, tetrahydrothien-3-yl, l-piperazinyl, and 2-piperazinyl.

[0254] In some embodiments, the conjugate moiety comprises an aryl or heteroaryl moiety. In some embodiments, the aryl group includes, among others, phenyl, naphthyl, indenyl, biphenyl, phenanthrenyl, naphthacenyl, anthracenyl, fluorenyl, indenyl, and azulenyl. In some embodiments, a heteroaryl group includes, among others, pyridyl, furanyl, thienyl, pynolyl, oxazolyl, oxadiazolyl,

imidazolyl ihiazolyl, isoxazolyl, quinolinyl, pyrazolyl, isoihiazolyl, pyridazinyl, pyrimidinyl, pyrazinyl, triazinyl, isoquinolinyl, and indazolyl.

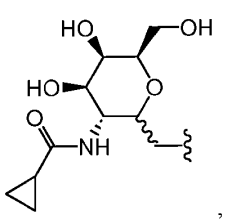
[0255] In some embodiments, the conjugate moiety comprises a cycloalkylalkyl-, heterocycloalkylalkyl-, arylalkyl-, heteroarylalkyl-, cycloalkylheteroalkyl-heterocycloalkylheteroalkyl-, arylheteroalkyl-, heteroarylheteroalkyl-, cycloalkylalkenyl-, heterocycloalkylalkenyl-, arylalkenyl-, heteroarylalkenyl-, cycloalkylheteroalkenyl-heterocycloalkylheteroalkenyl-, arylheteroalkenyl-, or heteroarylheteroalkenyl-.

[0256] In some embodiments, the conjugate moiety comprises a lipid or lipophilic moiety, for example a fatty acid. In some embodiments, the fatty acid comprises a saturated fatty acid, unsaturated fatty acid, or a polyunsaturated fatty acid. In some embodiments, the fatty acid comprises caprylic acid, lauric acid, myristic acid, palmitic acid, stearic acid, arachidic acid, behenic acid, oleic acid, elaidic acid, cis-vaccenic acid, trans-vaccenic acid, linoleic acid, alpha-linoleic acid, gamma-linoleic acid, arachidonic acid, eicosapentaenoic acid, decanoic acid, docosahexaenoic acid (DHA), and docosanoic acid (DCA) conjugate moieties (see, e.g., Kubo et al., ACS Chem. Biol., 2021, 16, 150−164; see also, WO2024/040041; incorporated herein by reference).

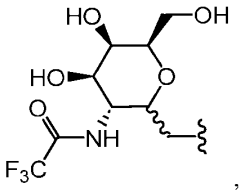
[0257] In some embodiments, the conjugate moiety comprises a sterol. In some embodiments, the sterol comprises cholesterol, alpha-cholesterol, cholesterol ester (e.g., cholesteryl palmitate, etc.), cholesterol sulfate, phytosterol, cholic acid, or lithocholic acid.

[0258] In some embodiments, the conjugate moiety comprises a phospholipid. In some embodiments, the phospholipid comprises phosphatidic acid, phosphatidylethanolamine, phosphatidylcholine, phosphatidylinositol, phosphatidylserine, or a sphingolipid.

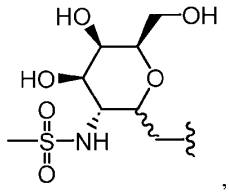
[0259] In some embodiments, the conjugate moiety comprises a carbohydrate, particularly a carbohydrate moiety acting as a ligand for a cellular receptor for cellular targeting of the oligonucleotide. In some embodiments, the carbohydrate moiety comprises galactose or galactose derivatives. In some embodiments, the carbohydrate moiety is attached to the nucleoside via a linker. In some embodiments, exemplary carbohydrates that can be used include the following.



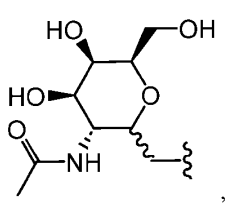
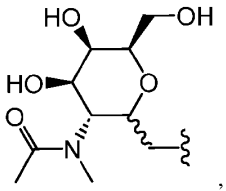
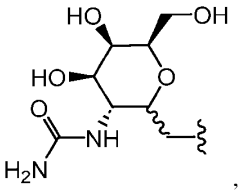
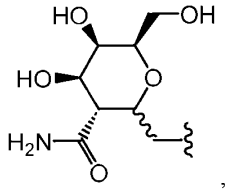
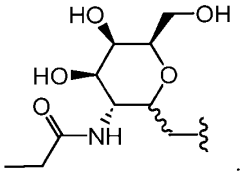
,



,

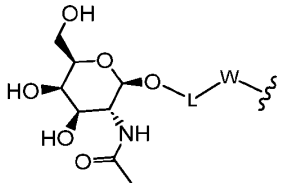


,

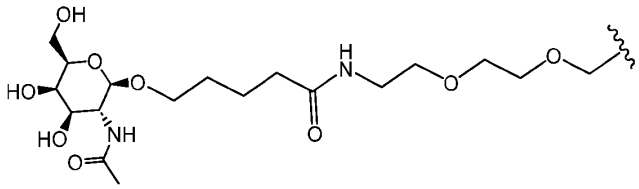






[0260] In some embodiments, the conjugate moiety is an N-acetylgalactosamine (GalNAc) conjugate moiety. In some embodiments, the oligonucleotide acceptor and/or nucleotide donor may be conjugated to at least one conjugate moiety comprising at least one N-acetylgalactosamine (GalNAc) moiety. In some embodiments, the conjugate moiety is a monovalent, divalent, trivalent or tetravalent, GalNAc.

[0261] In some embodiments, the GalNAc moiety has the following structure,

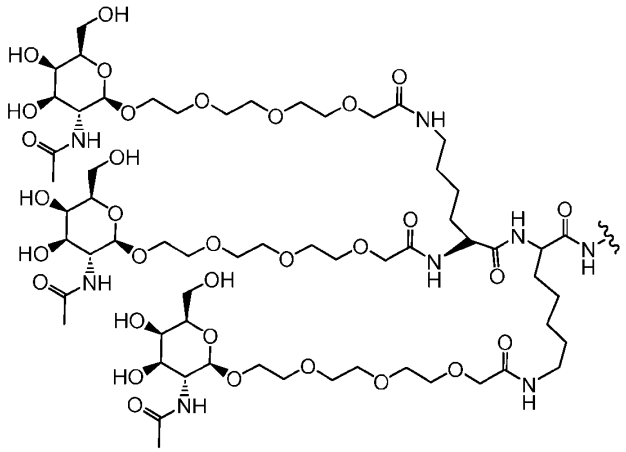


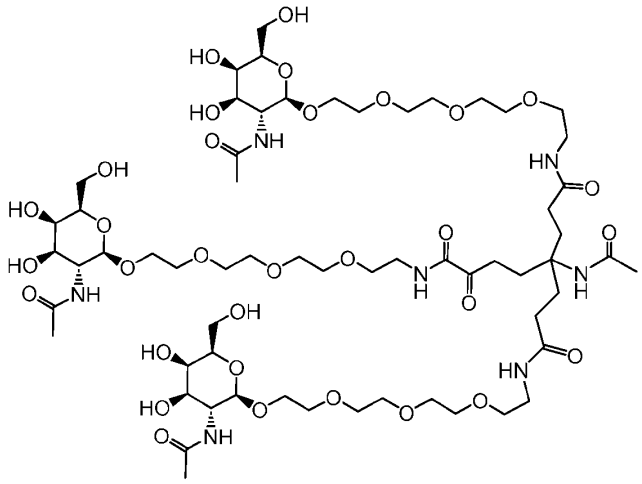
where L is a linker, and W is a heteroatom (e.g., O or N). In some embodiments, the W is the 2'-OH of the sugar moiety of a nucleoside. An exemplary monovalent GalNAc moiety is

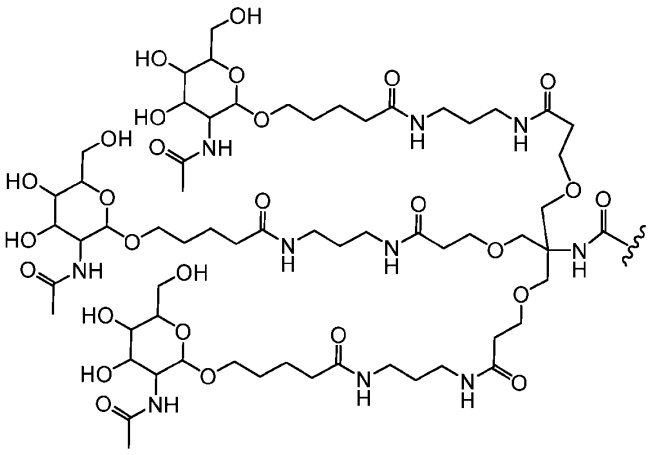


wherein the monovalent GalNAc is attached via the linker to the 2'-position of a nucleoside, such as adenine or guanine. These conjugate moieties can be present in contiguous nucleotides in an oligonucleotide (see, e.g., WO2024/040041).

[0262] In some embodiments, the conjugate moiety is a trivalent GalNAc. Tri-valent N-acetylgalactosamine conjugate moieties are described in, for example, WO 2014/076196, WO 2014/207232 and WO 2014/179620. The term "trivalent GalNAc" refers to a residue comprising three N-acetylgalactosamine moieties, typically attached via a linker. Exemplary trivalent GalNAc conjugate moieties are depicted below:







[0263] In some embodiments, the conjugate moiety comprises a reporter molecule. Examples of reporter molecules include, among others, fluorescent moieties, such as fluorescein and fluorescein dyes (e.g., fluorescein isothiocyanine or FITC, naphthofluorescein, 4′,5′-dichloro-2′,7′-dimethoxy-fluorescein, 6-carboxyfluorescein or FAM), carbocyanine, merocyanine, styryl dyes, oxonol dyes, phycoerythrin, erythrosin, eosin, rhodamine dyes (e.g., carboxytetramethylrhodamine or TAMRA, carboxyrhodamine 6G, carboxy-X-rhodamine (ROX), lissamine rhodamine B, rhodamine 6G, rhodamine Green, rhodamine Red, tetramethylrhodamine or TMR), coumarin and coumarin dyes (e.g., methoxycoumarin, dialkylaminocoumarin, hydroxycoumarin and aminomethylcoumarin or AMCA), Oregon Green Dyes (e.g., Oregon Green 488, Oregon Green 500, Oregon Green 514), Texas Red, Texas Red-X, Spectrum Red™, Spectrum Green™, cyanine dyes (e.g., Cy-3™, Cy-5™, Cy-3.5™, Cy-5.5™), Alexa Fluor dyes (e.g., Alexa Fluor 350, Alexa Fluor 488, Alexa Fluor 532, Alexa Fluor 546, Alexa Fluor 568, Alexa Fluor 594, Alexa Fluor 633, Alexa Fluor 660 and Alexa Fluor 680), BODIPY dyes (e.g., BODIPY FL, BODIPY R6G, BODIPY TMR, BODIPY TR, BODIPY 530/550, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY 630/650, BODIPY 650/665), IRDyes (e.g., IRD40, IRD 700, IRD 800). (See, e.g., "The Handbook of Fluorescent Probes and Research Products", 9th Ed., R.P. Haugland, 2002, Molecular Probes, Inc., Eugene, Oregon)

[0264] In some embodiments, the reporter moiety is a chemiluminescent moiety, for example acridinium esters, ruthenium derivatives (e.g., tris(2,2′-bipyridyl) ruthenium), and dioxetanes.

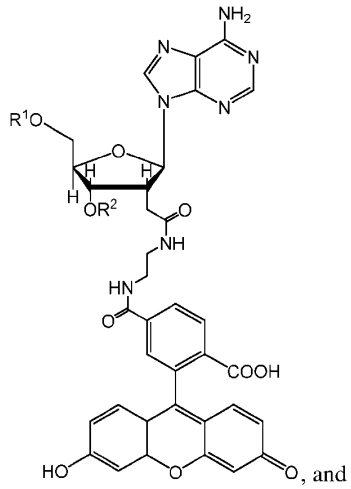
[0265] In some embodiments, the conjugate moiety comprises an affinity or capture tag. Exemplary affinity or capture tag includes, among others, biotin, desthiobiotin, digoxigenin, 3-amino-3-deoxydigoxigenin, and a hapten (e.g., dinitrophenol, Alexa Fluor 40, Alexa Fluor 488, dansyl, Lucifer yellow, Oregon Green 488, fluorescein).

[0266] In some embodiments, the conjugate moiety comprises a peptide. In some embodiments, the peptide comprises a cellular targeting peptide and/or cell penetration peptide (CPP) for enhancing cellular delivery of a conjugate modified oligonucleotide. In some embodiments, the cell penetrating peptide is attached via a linker, including a cleavable linker. Cell penetrating peptides, include among others, TAT, penetratin, MAP, transportan/TP10, VP22, polyarginine, MPG, Pep-1, pVEC, YTA2, YTA4, M918, and CADY. In some embodiments, the conjugate moiety comprises an RGD (Arg-Gly-Asp) peptide. Sequence of some penetrating peptides are described in Copolovici et al., 2014, 8(3):1972–1994 and some are provided below:

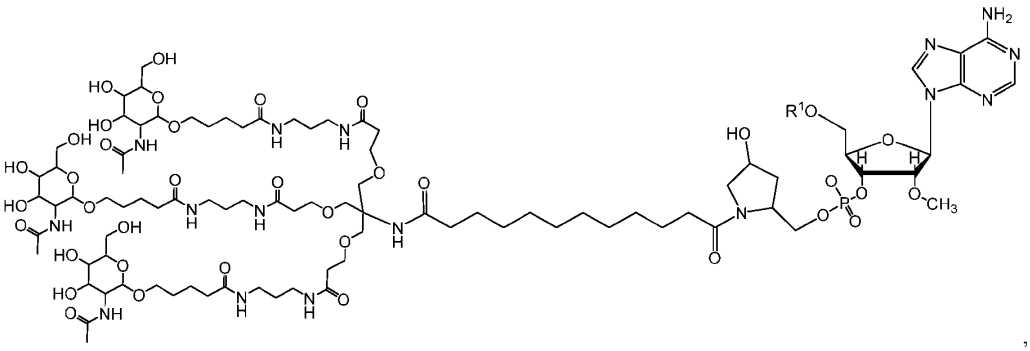
CPP Peptide	Peptide Sequence
TAT	GRKKRRQRRRPPQ
penetratin	RQIKIWFQNRRMKWKK
MAP	KLALKLALKALKAALKLA
Transportan/TP10	GWTLNS/AGYLLGKINLKALAALAKKIL
Transsortan9/TP9	LKALLGKIN
VP22	NAKTRRHERRRKLAIER
Polyarginine	(R)n, where n is 8 or 9
MPG	GALFLGFLGAAGSTMGA
Pep-1	KETWWETWWTEWSQPKKKRKV
pVEC	LLIILRRRIRKQAHAHSK
YTA2,	YTAIAWVKAFIRKLRK
YTA4,	IAWVKAFIRKLRKGPLG
M918	MVTVLFRRLRIRRACGPPRVRV
CADY	GLWRALWRLLRSLWRLLWRA
RGD	RGD

[0267] Other cell penetrating peptides, including those conjugated to nucleic acids, are disclosed in, among others, patent publications WO24063570, WO24044663, US2024083949, WO24026141, WO23230600, WO23219933, WO23177261, WO23178327, WO23093960, WO23086342, WO23081893, WO23069332, WO23070108, WO23034515, US2023248630, US2023053924, WO23003380, WO23277628, WO23277575, US2022378946, WO22171972, WO22162200, WO2020144233, WO22180242, WO22132520, WO22129926, WO22125673, WO22120276, WO22101193, US2023287086, US2023357334, US2023144488, and US2023048338; incorporated by reference herein. In some embodiments, the peptide can be attached using a thiol group on the 5'-phosphate of a polynucleotide or oligonucleotide.

[0268] Exemplary nucleotides/nucleosides with conjugate moieties are shown below:



, and

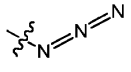


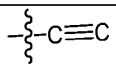
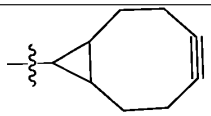
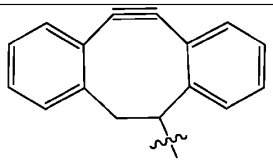
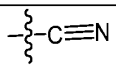
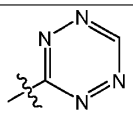
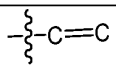
,

wherein R^1 is H or phosphate group, and R^2 is H, phosphate group, or blocking group.

Reactive Groups

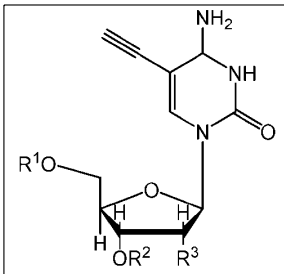
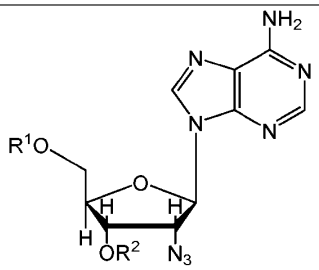
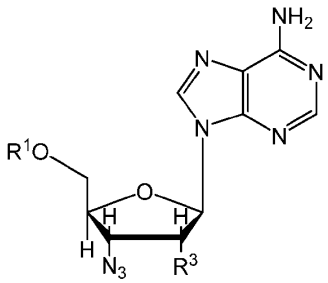
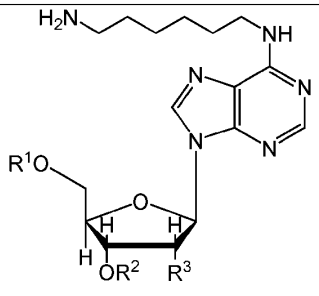
[0269] In some embodiments, the modification comprises a reactive group that is conjugated to a nucleoside. In some embodiments, the reactive group is attached to the nucleoside via a linker. In some embodiments, the reactive group is a cyano, azido, alkynyl, amino, carboxyl, sulfhydryl, dibenzocyclooctynyl, vinyl, trans-cyclooctene, or tetrazine. In some embodiments, the reactive group is used for click chemistry, including copper free click chemistry. Exemplary reactive groups are provided below:

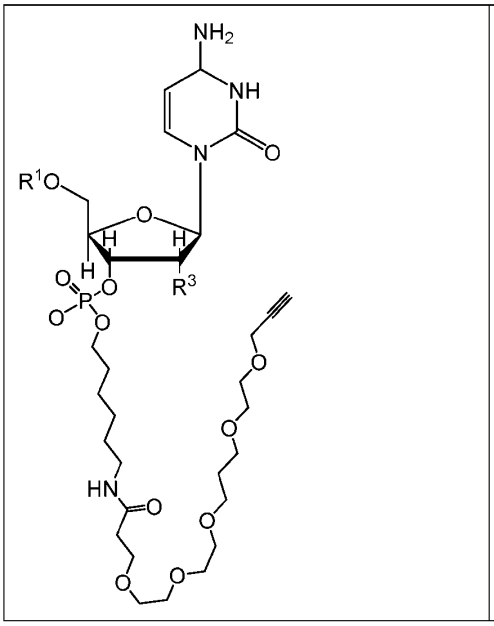
amino	
azido	

alkynyl	
BCN (bicyclo[6.1.0]nonyne)	
dibenzocyclooctynyl	
cyano	
tetrazinyl	
vinyl	

[0270] Other reactive groups used in click chemistry, particularly for nucleic acids, are described in Fantoni et al., Chem. Rev. 2021, 121, 7122−7154, incorporated by reference herein.

[0271] Exemplary substrates with a reactive moiety are provided below:



wherein R^1 is H or phosphate; R^2 is a blocking group or H; and R^3 is H, -OR, or halo, e.g., F, Br, or Cl.

Linkers

[0272] In some embodiments, as described above, the conjugate moiety or reactive moiety is attached to the nucleoside or the terminal group through a linker. Various linkers are known in the art for conjugating chemical groups to nucleosides and phosphate groups.

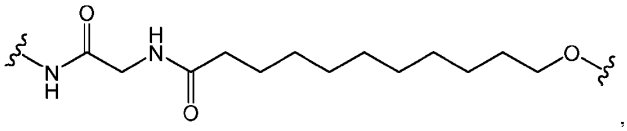
[0273] In some embodiments, mixtures of linkers are used. In some embodiments, different linker types are connected to form a longer linker or linkers with branched or dendritic structure. For example, an alkylene linker is connected to a polyethylene linker through a functional group, e.g., an amide; an arylene linker is attached to an alkylene linker. As such, different combinations of linker types can be connected to provide for longer linkers and/or branched linkers, for example for attaching multiple conjugate moieties.

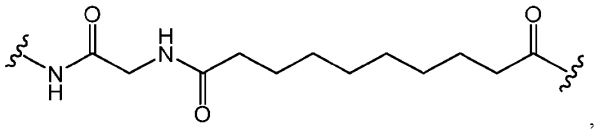
[0274] In some embodiments, linkers include, among others, substituted or unsubstituted alkylene, heteroalkylene, alkenylene, heteroalkenylene, arylene, heteroarylene, arylalkylene, arylalkenylene, heteroarylalkylene, heteroarylalkenylene, arylheteroalkylene, arylheteroalkenylene, heteroarylheteroalkylene, and heteroarylalkenylene. In some embodiments, the linker comprises substituted or unsubstituted C_2-C_{22} alkylene, heteroalkylene, or polyethylene glycol. In some embodiments, the linkers have functional groups for conjugation.

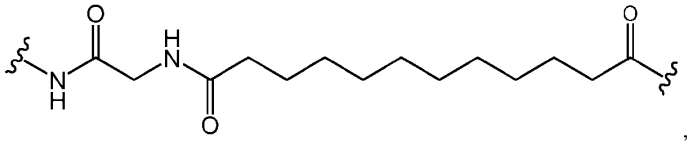
[0275] In some embodiments, the linker comprises a divalent, branched or unbranched, saturated or unsaturated, hydrocarbon chain, having from 1 to 50 carbon atoms, 1 to 20 carbon atoms, or 1 to 14 carbon atoms, wherein one or more of the carbon atoms in the hydrocarbon chain is optionally replaced by -O-, -NR^1-, -NR^1-C(=O)-, -C(=O)-NR^1, or -S-, and wherein R^1 is hydrogen or (C_1-C_6)alkyl, wherein the hydrocarbon chain, is optionally substituted with one or more (e.g. 1, 2, 3, or 4) substituents selected from (C_1-C_6)alkoxy, (C_3-C_6)cycloalkyl, (C_1-C_6)alkanoyl, (C_1-C_6)alkanoyloxy, (C_1-C_6)alkoxycarbonyl, (C_1-C_6)alkylthio, azido, cyano, nitro, halo, hydroxy, oxo (=O), carboxy, aryl, aryloxy, heteroaryl, and heteroaryloxy.

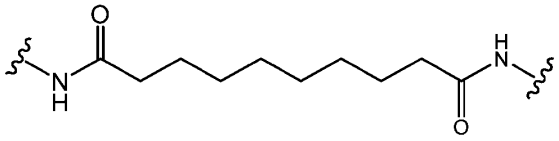
[0276] In some embodiments, the linker is attached to the nucleoside and/or conjugate through -NH-, -O-, -S-, -(C=O)-, -(C=O)-NH-, -NH-(C=O)-, -(C=O)-O-, -NH-(C=O)-NH-, or -NH-(SO_2)-.

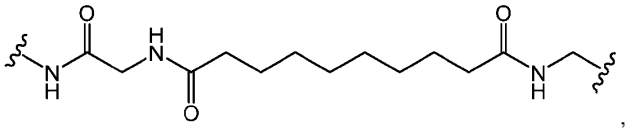
[0277] In some embodiments, the linker has the structure below:

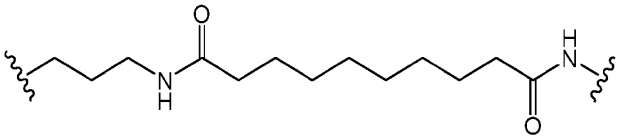

,

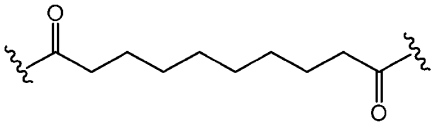

,


,

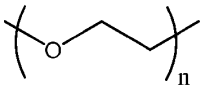

,


, or



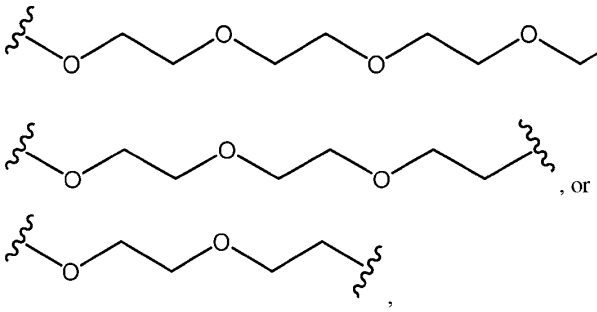


[0278] In some embodiments, the linker comprises a substituted or unsubstituted polyethylene glycol linker. In some embodiments, the polyethylene glycol linker has the formula:

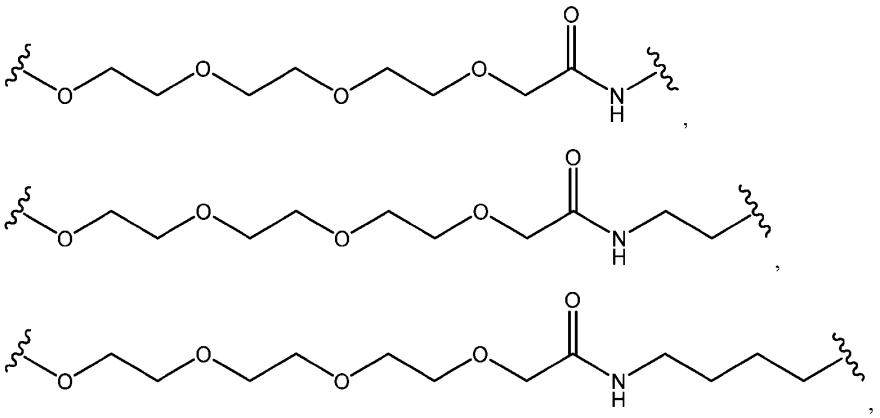


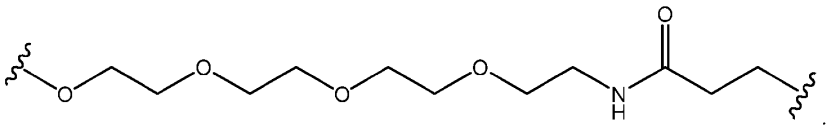
[0279] In some embodiments, n is 2-24. In some embodiments, n is 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, or 24.

[0280] In some embodiments, the polyethylene linker has the structure below:



[0281] In some embodiments, the polyethylene linker has the structure below:





[0282] In some embodiments, the linker is a cleavable linker in which the linker can be cleaved, for example to detach a conjugate moiety. Example of a cleavable linker includes, by way of example and not limitation, a disulfide linkage, enzymatically cleavable linkers (e.g., peptide linkers), and photocleavable linkers (see, e.g., Hermanson, G., Bioconjugate Techniques, 3rd Ed., 2013, Academic Press; see also Bioconjugation Protocols: Strategies and Methods, In Methods in Molecular Biology, 2nd Ed., S.S. Mark ed., 2011, Humana Press).

[0283] In some embodiments, bifunctional linkers can be used to attach a conjugate moiety to the linker and attach the linker-conjugate to the nucleoside or vice versa (see, e.g., Hermanson, G., supra; see also Bioconjugation Protocols: Strategies and Methods, In Methods in Molecular Biology, supra). In some embodiments, an activating group can be attached to an atom to activate the atom to form a covalent bond with another reactive group. Examples of synthetic activating groups that can be attached to an oxygen atom include, but are not limited to, acetate, succinate, triflate, and mesylate. When an activating group is attached to an oxygen atom of a carboxylic acid, the activating group can be a group that is derivable from a known coupling reagent. Examples of such coupling reagents include, but are not limited to, N,N′-dicyclohexylcarbodimide (DCC), hydroxybenzotriazole (HOBt), N-(3-dimethylaminopropyl)-N′-ethylcarbonate (EDC), (denzotriazol-1-yloxy)tris(dimethylamino)phosphonium hexafluorophosphate (BOP), benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate (PyBOP) or O-benzotriazol-1-yl-N,N,N′,N′-tetramethyluronium hexafluorophosphate (HBTU).

Specific Embodiments and Reaction Conditions

[0284] In some specific embodiments described in the Examples, the NTP substrate is NTP- N6-Propargyl-rATP, 2-Ethynyl-rATP, 5-Ethynyl-rUTP, 3'-(O-Propargyl)-rATP, 3'-Azido-3'-rATP, 2'-Azido-2'-dATP, N6-(6-Aminohexyl)-rATP, or EDA-rATP-Cy3. However, the present disclosure is not limited to these examples.

[0285] It is to be understood from the present Examples, that in some embodiments, the conjugate moiety, reactive group, or linker acts as a removable blocking group or RBG, preventing the addition of additional NTP-3'-O-RBGs or natural or modified NTP substrates (particularly when the conjugate moiety or reactive group is positioned at or blocks the 3' position of nucleotide sugar). Similarly, in some embodiments, the conjugate moiety, reactive group, or linker may be incorporated into the interior of a growing polynucleotide chain through subsequent rounds of NTP or NTP-RBG addition, as demonstrated in the Examples.

[0286] It is to be understood from the Examples that the TnTs of the present disclosure are capable of adding a NTP substate comprising a conjugate moiety, reactive group, or linker onto an oligo acceptor substrate comprising a conjugate moiety, reactive group, or linker. Specific examples of this include the incorporation of multiple iterative additions of N6-Propargyl-rATP, 2-Ethynyl-rATP, 5-Ethynyl-rUTP, 2'-Azido-2'-dATP, and N6-(6-Aminohexyl)-rATP in Example 3 (denoted as N+1, N+2, N+3, and >N+3 additions).

[0287] The substrate compound(s) in the reaction mixtures can be varied, taking into consideration, for example, the desired amount of product compound, the effect of each substrate concentration on enzyme activity, stability of enzyme under reaction conditions, and the percent conversion of each substrate to product. In some embodiments, the suitable reaction conditions comprise a substrate compound loading for each oligo acceptor substrate of at least about 0.1 μM to 1 μM, 1 μM to 2 μM, 2 μM to 3 μM, 3 μM to 5 μM, 5 μM to 10 μM, or 10 μM to 50 μM,, or 50 μM to 100 μM, or 100 μM to 500 μM, or 500 μM to 1000 μM, or 1000 μM to 2000 μM, or 2000 μM to 5000 μM, or 5000 μM to 10000, or 10000 μM or greater. In some embodiments, the suitable reaction conditions comprise a substrate compound loading for each oligo acceptor substrate comprising one or more conjugate moieties, reactive groups, or linkers of at least about 0.5 to about 25 g/L, 1 to about 25 g/L, 5 to about 25 g/L, about 10 to about 25 g/L, or 20 to about 25 g/L. In some embodiments, the suitable reaction conditions comprise a substrate compound loading for each oligo acceptor substrate of at least about 0.5 g/L, at least about 1 g/L, at least about 5 g/L, at least about 10 g/L, at least about 15 g/L, at least about 20 g/L, or at least about 30 g/L, or even greater.

[0288] In some embodiments, the suitable reaction conditions comprise a substrate compound loading for each NTP-3'-O-RBG or natural or modified NTP substrate comprising one or more conjugate moieties, reactive groups, or linkers of at least about 1 μM to 5 μM, 5 μM to 10 μM, 10 μM to 25 μM, 25 μM to 50 μM, 50 μM to 100 μM, 100 μM to 200 μM, 200 μM to 300 μM, 300 μM to 500 μM, 400 μM to 600 μM, 700 μM to 900 μM, 800 μM to 1000 μM, 800 μM to 1200 μM, or 1000 μM to 1500 μM, or 1600 μM to 2500 μM, or 2500 μM to 5000 μM, or 6000 μM to 12000 μM. In some embodiments, the suitable reaction conditions comprise a substrate compound loading for each NTP substrate of at least about 0.5 g/L, at least about 1 g/L, at least about 5 g/L, at least about 10 g/L, at least about 15 g/L, at least about 20 g/L, or at least about 30 g/L, or even greater.

[0289] In carrying out the TnT-mediated synthesis processes described herein, the engineered polypeptide may be added to the reaction mixture in the form of a purified enzyme, partially purified enzyme, whole cells transformed with gene(s) encoding the enzyme, as cell extracts and/or lysates of such cells, and/or as an enzyme immobilized on a solid support. Whole cells transformed with gene(s) encoding the engineered TnT enzyme or cell extracts, lysates thereof, and isolated enzymes may be employed in a variety of different forms, including solid (e.g., lyophilized, spray-dried, and the like) or semisolid (e.g., a crude paste). The cell extracts or cell lysates may be partially purified by

precipitation (ammonium sulfate, polyethyleneimine, heat treatment or the like, followed by a desalting procedure prior to lyophilization (e.g., ultrafiltration, dialysis, etc.). Any of the enzyme preparations (including whole cell preparations) may be stabilized by crosslinking using known crosslinking agents, such as, for example, glutaraldehyde or immobilization to a solid phase (e.g., Eupergit C, and the like).

[0290] The gene(s) encoding the engineered TnT polypeptides can be transformed into host cell separately or together into the same host cell. For example, in some embodiments one set of host cells can be transformed with gene(s) encoding one engineered TnT polypeptide, and another set can be transformed with gene(s) encoding another TnT. Both sets of transformed cells can be utilized together in the reaction mixture in the form of whole cells, or in the form of lysates or extracts derived therefrom. In other embodiments, a host cell can be transformed with gene(s) encoding multiple engineered TnT polypeptides. In some embodiments the engineered polypeptides can be expressed in the form of secreted polypeptides, and the culture medium containing the secreted polypeptides can be used for the TnT reaction.

[0291] In some embodiments, the improved activity of the engineered TnT polypeptides disclosed herein provides for processes wherein higher percentage conversion can be achieved with lower concentrations of the engineered polypeptide. In some embodiments of the process, the suitable reaction conditions comprise an engineered polypeptide amount of about 1% (w/w), 2% (w/w), 5% (w/w), 10% (w/w), 20% (w/w), 30% (w/w), 40% (w/w), 50% (w/w), 75% (w/w), 100% (w/w) or more of substrate compound loading.

[0292] In some embodiments, the engineered polypeptide is present at a molar ratio of engineered polypeptide to substrate of about 50 to 1, 25 to 1, 10 to 1, 5 to 1, 1 to 1, 1 to 5, 1 to 10, 1 to 25 or 1 to 50, 1 to 100, 1 to 500, 1 to 1000, or 1 to 2000. In some embodiments, the engineered polypeptide is present at a molar ratio of engineered polypeptide to substrate from a range of about 50 to 1 to a range of about 1 to 2000.

[0293] In some embodiments, the suitable reaction conditions comprise a divalent metal cofactor. In some embodiments, the divalent metal cofactor is cobalt. In some embodiments, the cobalt (II) chloride is present at concentrations of about 1 to 1500 μM, about 50 to 400 μM, about 100 to 300 μM, about 200 to 600 μM, or about 500 to 1000 μM. In some embodiments, the cobalt (II) chloride is present at concentrations of about 150 μM, about 200 μM, about 250 μM, about 500 μM, 1000 μM, or about 1500 μM.

[0294] In some embodiments, the engineered polypeptide is present at about 0.01 g/L to about 50 g/L; about 0.01 to about 0.1 g/L; about 0.05 g/L to about 50 g/L; about 0.1 g/L to about 40 g/L; about 1 g/L to about 40 g/L; about 2 g/L to about 40 g/L; about 5 g/L to about 40 g/L; about 5 g/L to about 30 g/L; about 0.1 g/L to about 10 g/L; about 0.5 g/L to about 10 g/L; about 1 g/L to about 10 g/L;

about 0.1 g/L to about 5 g/L; about 0.5 g/L to about 5 g/L; or about 0.1 g/L to about 2 g/L. In some embodiments, the TnT polypeptide is present at about 0.01 g/L, 0.05 g/L, 0.1 g/L, 0.2 g/L, 0.5 g/L, 1, 2 g/L, 5 g/L, 10 g/L, 15 g/L, 20 g/L, 25 g/L, 30 g/L, 35 g/L, 40 g/L, or 50 g/L.

[0295] In some embodiments of the reaction, a phosphatase is used to degrade inorganic phosphate and shift the reaction equilibrium toward the oligo acceptor extension product. In some embodiments, the phosphatase is an *E. coli* pyrophosphatase. In some embodiments, the phosphatase is from *Methanotorris igneus*. In some embodiments, the phosphatase is the phosphatase of SEQ ID NO: 2. In some embodiments, the phosphatase is present at a concentration of about 0.0001 to 0.01 units/uL; about 0.001 to 0.005 units/uL; or about 0.002 to 0.003 units/uL. In some embodiments, the phosphatase is present at a concentration of about 0.001 units/uL; about 0.002 units/uL; or about 0.003 units/uL. In some embodiments, the phosphatase is from *Geobacillus zalihae*. In some embodiments, the phosphatase is present at a concentration of about 0.01 to 10 μM; about 0.01 to 0.1 μM; or about 0.1 to 1 μM; or about 0.1 to 10 μM. In some embodiments, the phosphatase is present at a concentration of about 0.05 μM; about 0.5 μM; or about 5 μM; or about 10 μM.

[0296] During the course of the reaction, the pH of the reaction mixture may change. The pH of the reaction mixture may be maintained at a desired pH or within a desired pH range. This may be done by the addition of an acid or a base, before and/or during the course of the reaction. Alternatively, the pH may be controlled by using a buffer. Accordingly, in some embodiments, the reaction condition comprises a buffer. Suitable buffers to maintain desired pH ranges are known in the art and include, by way of example and not limitation, borate, phosphate, 2-(N-morpholino)ethanesulfonic acid (MES), 3-(N-morpholino)propanesulfonic acid (MOPS), acetate, triethanolamine (TEoA), and 2-amino-2-hydroxymethyl-propane-1,3-diol (Tris), and the like. In some embodiments, the reaction conditions comprise water as a suitable solvent with no buffer present.

[0297] In the embodiments of the process, the reaction conditions comprise a suitable pH. The desired pH or desired pH range can be maintained by use of an acid or base, an appropriate buffer, or a combination of buffering and acid or base addition. The pH of the reaction mixture can be controlled before and/or during the course of the reaction. In some embodiments, the suitable reaction conditions comprise a solution pH from about 4 to about 10, pH from about 5 to about 10, pH from about 5 to about 9, pH from about 6 to about 9, pH from about 6 to about 8. In some embodiments, the reaction conditions comprise a solution pH of about 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5, or 10.

[0298] In the embodiments of the processes herein, a suitable temperature is used for the reaction conditions, for example, taking into consideration the increase in reaction rate at higher temperatures, and the activity of the enzyme during the reaction time period. Accordingly, in some embodiments, the suitable reaction conditions comprise a temperature of about 10 °C to about 95 °C, about 10 °C to

about 75 °C, about 15 °C to about 95 °C, about 20 °C to about 95 °C, about 20 °C to about 65 °C, about 25 °C to about 70 °C, or about 50 °C to about 70 °C. In some embodiments, the suitable reaction conditions comprise a temperature of about 10 °C, 15 °C, 20 °C, 25 °C, 30 °C, 35 °C, 40 °C, 45 °C, 50 °C, 55 °C, 60 °C, 65 °C, 70 °C, 75 °C, 80 °C, 85 °C, 90 °C or 95 °C. In some embodiments, the temperature during the enzymatic reaction can be maintained at a specific temperature throughout the course of the reaction. In some embodiments, the temperature during the enzymatic reaction can be adjusted over a temperature profile during the course of the reaction.

[0299] In some embodiments, the processes of the invention are carried out in a solvent. Suitable solvents include water, aqueous buffer solutions, organic solvents, polymeric solvents, and/or co-solvent systems, which generally comprise aqueous solvents, organic solvents and/or polymeric solvents. The aqueous solvent (water or aqueous co-solvent system) may be pH-buffered or unbuffered. In some embodiments, the processes using the engineered TnT polypeptides can be carried out in an aqueous co-solvent system comprising an organic solvent (e.g., ethanol, isopropanol (IPA), dimethyl sulfoxide (DMSO), dimethylformamide (DMF) ethyl acetate, butyl acetate, 1-octanol, heptane, octane, methyl t butyl ether (MTBE), toluene, and the like), ionic or polar solvents (e.g., 1-ethyl 4 methylimidazolium tetrafluoroborate, 1-butyl-3-methylimidazolium tetrafluoroborate, 1-butyl 3 methylimidazolium hexafluorophosphate, glycerol, polyethylene glycols, and the like). In some embodiments, the co-solvent can be a polar solvent, such as a polyol, dimethylsulfoxide (DMSO), or lower alcohol. The non-aqueous co- solvent component of an aqueous co-solvent system may be miscible with the aqueous component, providing a single liquid phase, or may be partly miscible or immiscible with the aqueous component, providing two liquid phases. Exemplary aqueous co-solvent systems can comprise water and one or more co-solvents selected from an organic solvent, polar solvent, and polyol solvent. In general, the co-solvent component of an aqueous co-solvent system is chosen such that it does not adversely inactivate the TnT enzyme under the reaction conditions. Appropriate co-solvent systems can be readily identified by measuring the enzymatic activity of the specified engineered TnT enzyme with a defined substrate of interest in the candidate solvent system, utilizing an enzyme activity assay, such as those described herein.

[0300] In some embodiments of the process, the suitable reaction conditions comprise an aqueous co-solvent, where the co-solvent comprises DMSO at about 1% to about 50% (v/v), about 1 to about 40% (v/v), about 2% to about 40% (v/v), about 5% to about 30% (v/v), about 10% to about 30% (v/v), or about 10% to about 20% (v/v). In some embodiments of the process, the suitable reaction conditions can comprise an aqueous co-solvent comprising ethanol at about 1% (v/v), about 5% (v/v), about 10% (v/v), about 15% (v/v), about 20% (v/v), about 25% (v/v), about 30% (v/v), about 35% (v/v), about 40% (v/v), about 45% (v/v), or about 50% (v/v).

[0301] In some embodiments, the reaction conditions comprise a surfactant for stabilizing or enhancing the reaction. Surfactants can comprise non-ionic, cationic, anionic and/or amphiphilic

surfactants. Exemplary surfactants, include by way of example and not limitation, nonyl phenoxypolyethoxylethanol (NP40), TRITON™ X-100 polyethylene glycol *tert*-octylphenyl ether, polyoxyethylene-stearylamine, cetyltrimethylammonium bromide, sodium oleylamidosulfate, polyoxyethylene-sorbitanmonostearate, hexadecyldimethylamine, etc. Any surfactant that may stabilize or enhance the reaction may be employed. The concentration of the surfactant to be employed in the reaction may be generally from 0.1 to 50 mg/mL, particularly from 1 to 20 mg/mL.

[0302] In some embodiments, the reaction conditions include an antifoam agent, which aids in reducing or preventing formation of foam in the reaction solution, such as when the reaction solutions are mixed or sparged. Anti-foam agents include non-polar oils (e.g., minerals, silicones, etc.), polar oils (e.g., fatty acids, alkyl amines, alkyl amides, alkyl sulfates, etc.), and hydrophobic (e.g., treated silica, polypropylene, etc.), some of which also function as surfactants. Exemplary anti-foam agents include Y-30® (Dow Corning), poly-glycol copolymers, oxy/ethoxylated alcohols, and polydimethylsiloxanes. In some embodiments, the anti-foam can be present at about 0.001% (v/v) to about 5% (v/v), about 0.01% (v/v) to about 5% (v/v), about 0.1% (v/v) to about 5% (v/v), or about 0.1% (v/v) to about 2% (v/v). In some embodiments, the anti-foam agent can be present at about 0.001% (v/v), about 0.01% (v/v), about 0.1% (v/v), about 0.5% (v/v), about 1% (v/v), about 2% (v/v), about 3% (v/v), about 4% (v/v), or about 5% (v/v) or more as desirable to promote the reaction.

[0303] The quantities of reactants used in the TnT reaction will generally vary depending on the quantities of product desired, and concomitantly the amount of substrates employed. Those having ordinary skill in the art will readily understand how to vary these quantities to tailor them to the desired level of productivity and scale of production.

[0304] In some embodiments, the order of addition of reactants is not critical. The reactants may be added together at the same time to a solvent (e.g., monophasic solvent, biphasic aqueous co-solvent system, and the like), or alternatively, some of the reactants may be added separately, and some together at different time points. For example, the cofactor, co-substrate, and substrate may be added first to the solvent.

[0305] The solid reactants (e.g., enzyme, salts, etc.) may be provided to the reaction in a variety of different forms, including powder (e.g., lyophilized, spray dried, and the like), solution, emulsion, suspension, and the like. The reactants can be readily lyophilized or spray dried using methods and equipment that are known to those having ordinary skill in the art. For example, the protein solution can be frozen at -80°C in small aliquots, then added to a pre-chilled lyophilization chamber, followed by the application of a vacuum.

[0306] For improved mixing efficiency when an aqueous co-solvent system is used, the TnT, and co-substrate may be added and mixed into the aqueous phase first. The substrate may be added and mixed in, followed by the organic phase or the substrate may be dissolved in the organic phase and

mixed in. Alternatively, the substrate may be premixed in the organic phase, prior to addition to the aqueous phase.

[0307] The processes of the present disclosure are generally allowed to proceed until further conversion of substrate to product does not change significantly with reaction time (e.g., less than 10% of substrate being converted, or less than 5% of substrate being converted). In some embodiments, the reaction is allowed to proceed until there is complete or near complete conversion of substrate to product. Transformation of substrate to product can be monitored using known methods by detecting substrate and/or product, with or without derivatization. Suitable analytical methods include gas chromatography, HPLC, MS, and the like. In some embodiments, after suitable conversion to product, the reactants are separated from the oligo acceptor substrate extension product and additional reactants are added to the oligo acceptor substrate extension product to further extend the growing polynucleotide chain. The processes of the present disclosure may be used to iteratively extend the oligo acceptor extension product until a polynucleotide of a defined sequence and length is synthesized.

[0308] Any of the processes disclosed herein using the engineered polypeptides for the preparation of products can be carried out under a range of suitable reaction conditions, including but not limited to ranges of substrates, temperature, pH, solvent system, substrate loading, polypeptide loading, cofactor loading, and reaction time. In one example, the suitable reaction conditions comprise: (a) oligo acceptor substrate loading of about 0.1 – 5000 μM of substrate compound; (b) NTP-3'-O-RBG substrate or NTP loading of about 1 – 10000 μM of substrate compound; (c) of about 0.01 g/L to 5 g/L engineered polypeptide; (d) 100 to 5000 μM cobalt (II) chloride; (e) 5 to 100 mM triethanolamine buffer; (f) 0.05 to 10 μM pyrophosphatase; (g) pH at 5-9; and (h) temperature of about 15 °C to 70 °C. In some embodiments, the suitable reaction conditions comprise: (a) oligo acceptor substrate loading of about 400 μM of substrate compound; (b) NTP-3'-O-RBG or NTP substrate loading of about 200 μM of substrate compound; (c) of about 5 μM engineered polypeptide; (d) 1000 μM cobalt (II) chloride; (e) 100 mM triethanolamine buffer; (f) 1 μM pyrophosphatase; (g) pH at 7.8; and (h) temperature of about 50 °C. In some embodiments, the enzyme loading is between 1-30% w/w. In some embodiments, additional reaction components or additional techniques carried out to supplement the reaction conditions. These can include taking measures to stabilize or prevent inactivation of the enzyme, reduce product inhibition, shift reaction equilibrium to formation of the desired product.

[0309] In some embodiments, the present disclosure provides an engineered TnT, wherein said engineered TnT has improved activity on NTP-3'-RBGs or modified NTPs comprising one or more conjugate moieties, reactive groups, or linkers, such that said NTPs are incorporated with equivalent efficiency to native NTPs, as compared to another wild-type or engineered TnT. In some embodiments, the engineered TnT with improved activity, is an engineered TnT polypeptide

comprising an amino acid sequence having at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identity to SEQ ID NOs: 4, 6, 8, 10, and/or 12. In some embodiments, the present disclosure provides an engineered TnT, wherein said engineered TnT has improved activity on NTP-3'-RBGs or modified NTPs, such that said NTPs are incorporated onto a oligo acceptor substrate comprising one or more conjugate moieties, reactive groups, or linkers with equivalent efficiency to a native oligo acceptor substrate, as compared to another wild-type or engineered TnT. In some embodiments, the engineered TnT with improved activity, is an engineered TnT polypeptide comprising an amino acid sequence having at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identity to SEQ ID NOs: 4, 6, 8, 10, and/or 12.

Methods of Using Engineered TnTs for Incorporation of Conjugates

[0310] In some embodiments, the present disclosure provides a method for incorporation or addition of NTP-3'-RBGs or modified NTPs comprising one or more conjugate moieties, reactive groups, or linkers to the 3'-OH of an oligo acceptor substrate or growing polynucleotide chain. In some embodiments, the present disclosure provides a method for incorporation or addition of NTP-3'-RBGs or modified NTPs to the 3'-OH of an oligo acceptor substrate or growing polynucleotide chain comprising one or more conjugate moieties, reactive groups, or linkers.

[0311] In some embodiments, the present disclosure provides a method of adding NTP-3'-O-RBGs or natural or modified NTP substrates onto a polynucleotide or oligo acceptor substrate containing one or more conjugate moieties, reactive groups, or linkers. In some embodiments, the method further comprises an engineered TnT polypeptide comprising an amino acid sequence having at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence identity to a reference sequence of SEQ ID NOs: 4, 6, 8, 10, and/or 12 and comprising at least one substitution or one substitution set at one or more positions, wherein the positions are numbered with reference to SEQ ID NOs: 4, 6, 8, 10, and/or 12.

[0312] In some embodiments, the present disclosure provides a method of adding NTP-3'-O-RBGs or natural or modified NTP substrates onto a polynucleotide or oligo acceptor substrate containing one or more conjugate moieties, reactive groups, or linkers. In some embodiments, the method further comprises an engineered TnT polypeptide comprising an amino acid sequence having at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence identity to a reference sequence of SEQ ID NOs: 4, 6, 8, 10, and/or 12 and comprising at least one substitution or one substitution set at one or more positions, wherein the positions are numbered with reference to SEQ ID NOs: 4, 6, 8, 10, and/or 12.

[0313] In one embodiment, the present disclosure provides a method comprising: (a) providing at least one TnT polypeptide comprising an amino acid sequence having at least 80%, 85%, 90%, 91%,

92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence identity to a reference sequence of SEQ ID NOs: 4, 6, 8, 10, and/or 12 and comprising at least one substitution or one substitution set at one or more positions; (b) providing at least one oligo acceptor substrate, wherein the oligo acceptor substrate comprises a 3'-OH or equivalent; and (c) contacting the oligo acceptor substrate, the TnT or template-independent polymerase, and a nucleotide triphosphate, a modified nucleotide triphosphate, or a NTP-3'-O-RBG comprising one or more conjugate moieties, reactive groups, or linkers under conditions sufficient for the addition of the nucleotide, modified nucleotide, or nucleotide-3'-O-RBG to the 3' end. In some embodiments, step (c) optionally includes contacting the oligo acceptor substrate, the TnT or template-independent polymerase, and a nucleotide triphosphate, a modified nucleotide triphosphate, or NTP-3'-O-RBG comprising one or more conjugate moieties, reactive groups, or linkers with a phosphatase, such as an inorganic pyrophosphatase to convert pyrophosphate to inorganic phosphate. In another embodiment, the method further comprises (d) deblocking the oligonucleotide formed in step (c) at the protected 3'-O-position of the oligonucleotide product. In some embodiments, the method comprises (e) deactivating unreacted nucleotide triphosphates, modified nucleotide triphosphates, or NTP-3'-O-RBGs. In some embodiments, step (d) deblocking the oligonucleotide formed in step (c) at the protected 3'-O-position of the nucleotide-3'-O-RBG and step (e) deactivating unreacted nucleotide triphosphates, modified nucleotide triphosphates, or NTP-3'-O-RBGs occur simultaneously. In some embodiments, step (d) deblocking the oligonucleotide formed in step (c) at the protected 3'-O-position of the nucleotide-3'-O-RBG and step (e) deactivating unreacted nucleotide triphosphates, modified nucleotide triphosphates, or NTP-3'-O-RBGs occur simultaneously, wherein the NTP-3'-O-RBG comprises a 3' phosphate and a phosphatase is used to deblock the nucelotide-3'-O-RBG while simultaneously deactivating unreacted NTP-3'-O-RBGs by removing the 5' phosphates to leave nucleosides. In some embodiments, the method comprises an optional step (f) of removing excess nucleoside and/or excess inorganic phosphate and/or pyrophosphate from the reaction. In some embodiments, steps (a)-(c) or (a)-(d) or (a)-(e) or (a)-(f) are repeated until a desired oligonucleotide sequence is obtained. In another embodiment, the method further comprises (g) cleaving or releasing the growing or completed oligonucleotide chain from the oligo acceptor substrate.

[0314] In one embodiment, the present disclosure provides a method comprising: (a) providing at least one TnT polypeptide comprising an amino acid sequence having at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence identity to a reference sequence of SEQ ID NOs: 4, 6, 8, 10, and/or 12 and comprising at least one substitution or one substitution set at one or more positions; (b) providing at least one oligo acceptor substrate, wherein the oligo acceptor substrate comprises a 3'-OH or equivalent and wherein the oligo acceptor substrate comprises one or more conjugate moieties, reactive groups, or linkers; and (c) contacting the oligo acceptor substrate, the TnT or template-independent polymerase, and a nucleotide triphosphate, a modified nucleotide

triphosphate, or a NTP-3'-O-RBG under conditions sufficient for the addition of the nucleotide, modified nucleotide, or nucleotide-3'-O-RBG to the 3' end. In some embodiments, step (c) optionally includes contacting the oligo acceptor substrate, the TnT or template-independent polymerase, and a nucleotide triphosphate, a modified nucleotide triphosphate, or NTP-3'-O-RBG with a phosphatase, such as an inorganic pyrophosphatase to convert pyrophosphate to inorganic phosphate. In another embodiment, the method further comprises (d) deblocking the oligonucleotide formed in step (c) at the protected 3'-O-position of the oligonucleotide product. In some embodiments, the method comprises (e) deactivating unreacted nucleotide triphosphates, modified nucleotide triphosphates, or NTP-3'-O-RBGs. In some embodiments, step (d) deblocking the oligonucleotide formed in step (c) at the protected 3'-O-position of the nucleotide-3'-O-RBG and step (e) deactivating unreacted nucleotide triphosphates, modified nucleotide triphosphates, or NTP-3'-O-RBGs occur simultaneously. In some embodiments, step (d) deblocking the oligonucleotide formed in step (c) at the protected 3'-O-position of the nucleotide-3'-O-RBG and step (e) deactivating unreacted nucleotide triphosphates, modified nucleotide triphosphates, or NTP-3'-O-RBGs occur simultaneously, wherein the NTP-3'-O-RBG comprises a 3' phosphate and a phosphatase is used to deblock the nucelotide-3'-O-RBG while simultaneously deactivating unreacted NTP-3'-O-RBGs by removing the 5' phosphates to leave nucleosides. In some embodiments, the method comprises an optional step (f) of removing excess nucleoside and/or excess inorganic phosphate and/or pyrophosphate from the reaction. In some embodiments, steps (a)-(c) or (a)-(d) or (a)-(e) or (a)-(f) are repeated until a desired oligonucleotide sequence is obtained. In another embodiment, the method further comprises (g) cleaving or releasing the growing or completed oligonucleotide chain from the oligo acceptor substrate.

[0315] In either of the embodiments described above, the method may optionally comprise both i) an oligo acceptor substrate comprising one or more conjugate moieties, reactive groups, or linkers and ii) a nucleotide triphosphate, a modified nucleotide triphosphate, or a NTP-3'-O-RBG comprising one or more conjugate moieties, reactive groups, or linkers.

[0316] Any of methods described herein may comprise a nucleotide comprising a conjugate moiety, reactive group, or linker at the 3' end, 5' end, or at an internal position of the RNA or other polynucleotide. The RNA or other polynucleotide may comprise one or more conjugate moieties, reactive groups, or linkers at one or more nucleotide positions. Any of the methods described herein may optionally comprise another enzyme, such as a ligase.

[0317] In some embodiments of the described method, the oligo acceptor substrate and growing oligonucleotide chain are immobilized on a solid support. In some embodiments of the described method, the TnT or template-independent polymerase, oligo acceptor substrate and growing oligonucleotide chain are all in solution phase. The oligo substrate and growing oligo chain can be optionally substituted with a soluble tag that aids extended oligo product isolation and purification. In some embodiments of the described method, the TnT or template-independent polymerase is

immobilized. In some embodiments of the described method, the TnT or template-independent polymerase is simultaneously purified and immobilized on a solid support. In some embodiments, the immobilized TnT or template-independent polymerase is an engineered TnT with greater than 60% sequence identity to SEQ ID NOs: 4-12 and one or more substitutions or substitution sets in the amino acid sequence of the engineered TnT. In some embodiments, the immobilized TnT or template-independent polymerase is immobilized on a solid support.

[0318] In some embodiments, the engineered TnT polypeptides can be provided on a solid support, such as a membrane, resin, solid carrier, or other solid phase material. A solid support can be composed of organic polymers such as polystyrene, polyethylene, polypropylene, polyfluoroethylene, polyethyleneoxy, and polyacrylamide, as well as co-polymers and grafts thereof. A solid support can also be inorganic, such as glass, silica, controlled pore glass (CPG), reverse phase silica or metal, such as gold or platinum. The configuration of a solid support can be in the form of beads, spheres, particles, granules, a gel, a membrane or a surface. Surfaces can be planar, substantially planar, or non-planar. Solid supports can be porous or non-porous and can have swelling or non-swelling characteristics. A solid support can be configured in the form of a well, depression, or other container, vessel, feature, or location.

[0319] In some embodiments, the engineered TnT polypeptides of the present invention can be immobilized on a solid support such that they retain their improved activity, and/or other improved properties relative to the reference polypeptide of SEQ ID NOs: 4-12. In such embodiments, the immobilized polypeptides can facilitate the biocatalytic conversion of the substrate compounds or other suitable substrates to the product and after the reaction is complete are easily retained (e.g., by retaining beads on which polypeptide is immobilized) and then reused or recycled in subsequent reactions. Such immobilized enzyme processes allow for further efficiency and cost reduction. Accordingly, it is further contemplated that any of the methods of using the TnT polypeptides of the present invention can be carried out using the TnT polypeptides bound or immobilized on a solid support.

[0320] Methods of enzyme immobilization are well-known in the art. The engineered polypeptides can be bound non-covalently or covalently. Various methods for conjugation and immobilization of enzymes to solid supports (e.g., resins, membranes, beads, glass, etc.) are well known in the art (See e.g., Yi et al., Proc. Biochem., 42(5): 895-898 [2007]; Martin et al., Appl. Microbiol. Biotechnol., 76(4): 843-851 [2007]; Koszelewski et al., J. Mol. Cat. B: Enzymatic, 63: 39-44 [2010]; Truppo et al., Org. Proc. Res. Dev., published online: dx.doi.org/10.1021/op200157c; Hermanson, Bioconjugate Techniques, 2nd ed., Academic Press, Cambridge, MA [2008]; Mateo et al., Biotechnol. Prog., 18(3):629-34 [2002]; and "Bioconjugation Protocols: Strategies and Methods," *In* Methods in Molecular Biology, Niemeyer (ed.), Humana Press, New York, NY [2004]; the disclosures of each which are incorporated by reference herein). Solid supports useful for immobilizing the engineered

TnT of the present invention include but are not limited to beads or resins comprising polymethacrylate with epoxide functional groups, polymethacrylate with amino epoxide functional groups, CPG with epoxide functional groups, styrene/DVB copolymer or polymethacrylate with octadecyl functional groups. Exemplary solid supports useful for immobilizing the engineered TnT polypeptides of the present invention include, but are not limited to, EnginZyme (including, EziG-1, EziG-1, and EziG-3), chitosan beads, Eupergit C, and SEPABEADs (Mitsubishi) (including EC-EP, EC-HFA/S, EXA252, EXE119 and EXE120).

[0321] In some embodiments of the described method, the TnT or template-independent polymerase is immobilized, and the method comprises an aqueous liquid phase. In some embodiments, the method for template-independent synthesis of an oligonucleotide comprising an aqueous phase and an immobilized TnT or template-independent polymerase further comprises a column solid support. In some embodiments, the method for template-independent synthesis of an oligonucleotide comprising an aqueous phase and an immobilized TnT or template-independent polymerase further comprises a batch method with a solid support. In some embodiments, the oligo acceptor substrate and/or growing oligonucleotide chain are provided in an aqueous phase. In some embodiments, the nucleotide triphosphate, the modified nucleotide triphosphate, or the NTP-3'-O-RBG are provided in an aqueous phase. In some embodiments, the method further comprises removing unreacted nucleotide triphosphates, modified nucleotide triphosphates, or NTP-3'-O-RBGs from the oligo acceptor substrate and/or growing oligonucleotide chain. In some embodiments of the described method, the oligo acceptor substrate and oligonucleotide are immobilized. In some embodiments, neither the oligo acceptor substrate nor the TnT of template-independent polymerase are immobilized.

[0322] In some further embodiments, the method comprising an immobilized TnT or template-independent polymerase and an aqueous liquid phase further comprises the steps of (a) providing at least one TnT on a solid support, wherein said TnT comprises a polypeptide sequence comprising at least 60% identity to any of the even-numbered sequences of SEQ ID NOs: 4-12 and at least one substitution or substitution set in said polypeptide sequence as compared to a reference sequence of any of the even-numbered sequences of SEQ ID NOs: 4-12; (b) providing at least one oligo acceptor substrate in an aqueous phase, wherein the oligo acceptor substrate comprises a 3'-OH or equivalent; and (c) contacting the oligo acceptor substrate, the TnT or template-independent polymerase, and a nucleotide triphosphate, a modified nucleotide triphosphate, or NTP-3'-O-RBG comprising one or more conjugate moieties, reactive groups, or linkers under aqueous conditions sufficient for the addition of the nucleotide, modified nucleotide, or nucleotide-3'-O-RBG to the 3' end. In some embodiments, step (c) optionally includes contacting the oligo acceptor substrate, the TnT or template-independent polymerase, and a nucleotide triphosphate, a modified nucleotide triphosphate, or NTP-3'-O-RBG with a phosphatase, such as a pyrophosphatase to convert pyrophosphate to inorganic phosphate. In another embodiment, the method further comprises (d) deblocking the

oligonucleotide formed in step (c) at the protected 3'-O-position of the oligonucleotide product. In another embodiment, the method further comprises (e) deactivating the unreacted nucleotide triphosphates, modified nucleotide triphosphates, or NTP-3'-O-RBGs. In some embodiments, step (d) deblocking the oligonucleotide formed in step (c) at the protected 3'-O-position of the oligonucleotide product and step (e) deactivating unreacted nucleotide triphosphates, modified nucleotide triphosphates, or NTP-3'-O-RBGs occur simultaneously. In some embodiments, step (d) deblocking the oligonucleotide formed in step (c) at the protected 3'-O-position of the oligonucleotide product and step (e) deactivating unreacted nucleotide triphosphates, modified nucleotide triphosphates, or NTP-3'-O-RBGs occur simultaneously, wherein the NTP-3'-O-RBG comprises a 3' phosphate and a phosphatase is used to deblock the nucleotide-3'-O-RBG while simultaneously deactivating unreacted NTP-3'-O-RBGs by removing the 5' phosphates to leave nucleosides. In some embodiments, the method comprises an optional step (f) of removing excess nucleoside and/or excess inorganic phosphate and/or pyrophosphate from the reaction. In some embodiments, steps (a)-(c) or (a)-(d) or (a)-(e) or (a)-(f) are repeated until a desired oligonucleotide sequence is obtained. In another embodiment, the method further comprises (g) cleaving or releasing the growing or completed oligonucleotide chain from the oligo acceptor substrate once a desired nucleotide sequence is obtained. In some embodiments, any of steps (a)-(g) are completed on a solid support. In some embodiments the solid support is a column. In some embodiments, any of steps (a)-(g) are completed in an aqueous phase passing over one or a series of in line columns. In some embodiments, the solid support is used in a batch method.

[0323] In some further embodiments, the method comprising an immobilized TnT or template-independent polymerase and an aqueous liquid phase further comprises the steps of (a) providing at least one TnT on a solid support, wherein said TnT comprises a polypeptide sequence comprising at least 60% identity to any of the even-numbered sequences of SEQ ID NOs: 4-12 and at least one substitution or substitution set in said polypeptide sequence as compared to a reference sequence of any of the even-numbered sequences of SEQ ID NOs: 4-12; (b) providing at least one oligo acceptor substrate in an aqueous phase, wherein the oligo acceptor substrate comprises a 3'-OH or equivalent and wherein the oligo acceptor substrate comprises one or more conjugate moieties, reactive groups, or linkers; and (c) contacting the oligo acceptor substrate, the TnT or template-independent polymerase, and a nucleotide triphosphate, a modified nucleotide triphosphate, or NTP-3'-O-RBG under aqueous conditions sufficient for the addition of the nucleotide, modified nucleotide, or nucleotide-3'-O-RBG to the 3' end. In some embodiments, step (c) optionally includes contacting the oligo acceptor substrate, the TnT or template-independent polymerase, and a nucleotide triphosphate, a modified nucleotide triphosphate, or NTP-3'-O-RBG with a phosphatase, such as a pyrophosphatase to convert pyrophosphate to inorganic phosphate. In another embodiment, the method further comprises (d) deblocking the oligonucleotide formed in step (c) at the protected 3'-O-position of the

oligonucleotide product. In another embodiment, the method further comprises (e) deactivating the unreacted nucleotide triphosphates, modified nucleotide triphosphates, or NTP-3'-O-RBGs. In some embodiments, step (d) deblocking the oligonucleotide formed in step (c) at the protected 3'-O-position of the oligonucleotide product and step (e) deactivating unreacted nucleotide triphosphates, modified nucleotide triphosphates, or NTP-3'-O-RBGs occur simultaneously. In some embodiments, step (d) deblocking the oligonucleotide formed in step (c) at the protected 3'-O-position of the oligonucleotide product and step (e) deactivating unreacted nucleotide triphosphates, modified nucleotide triphosphates, or NTP-3'-O-RBGs occur simultaneously, wherein the NTP-3'-O-RBG comprises a 3' phosphate and a phosphatase is used to deblock the nucleotide-3'-O-RBG while simultaneously deactivating unreacted NTP-3'-O-RBGs by removing the 5' phosphates to leave nucleosides. In some embodiments, the method comprises an optional step (f) of removing excess nucleoside and/or excess inorganic phosphate and/or pyrophosphate from the reaction. In some embodiments, steps (a)-(c) or (a)-(d) or (a)-(e) or (a)-(f) are repeated until a desired oligonucleotide sequence is obtained. In another embodiment, the method further comprises (g) cleaving or releasing the growing or completed oligonucleotide chain from the oligo acceptor substrate once a desired nucleotide sequence is obtained. In some embodiments, any of steps (a)-(g) are completed on a solid support. In some embodiments the solid support is a column. In some embodiments, any of steps (a)-(g) are completed in an aqueous phase passing over one or a series of in line columns. In some embodiments, the solid support is used in a batch method.

[0324] In some embodiments, the method comprises optionally contacting the oligo acceptor substrate, the TnT or template-independent polymerase, and a nucleotide triphosphate, a modified nucleotide triphosphate, or NTP-3'-O-RBG with a phosphatase to convert pyrophosphate to inorganic phosphate. In some embodiments, the production of inorganic phosphate from pyrophosphate drives the extension reaction toward the N+1 product. In some embodiments, the phosphatase is an inorganic pyrophosphatase. In some embodiments, the inorganic pyrophosphatase is derived from *Methanotorris igneus, Thermocrinis ruber, Aquifex pyrophilus, Thermus oshimai, Sulfolobus sp. A20, Geobacillus zalihae, Bacillus thermozeamaize*, or *Bacillus smithii.* In some embodiments, the inorganic pyrophosphatase comprises SEQ ID NO: 2. In some embodiments, the inorganic pyrophosphatase may be immobilized on a solid support.

[0325] In some embodiments, the method for template-independent synthesis of an oligonucleotide comprises a of step deblocking the oligonucleotide at the protected 3'-O-position of the oligonucleotide product and/or a step of deactivating unreacted nucleotide triphosphates, modified nucleotide triphosphates, or NTP-3'-O-RBGs to nucleosides using a phosphatase. In some embodiments, the phosphatase is an alkaline phosphatase. In some embodiments, the alkaline phosphatase is derived from *Pyrococcus furiosus, Thermotoga maritima, Thermotoga sp. 50_64,*

Pseudothermotoga lettingae, *Thermotoga neapolitana*, *Thermoflexibacter ruber*, or *Bacillus licheniformis*. In some embodiments, the alkaline phosphatase may be immobilized on a solid support.

[0326] In some embodiments, the method for template-independent synthesis of an oligonucleotide comprises an optional step of removing excess inorganic phosphate or nucleoside from the reaction.

[0327] In some embodiments, the method for template-independent synthesis of an oligonucleotide comprises an optional step of cleaving or releasing the growing or completed oligonucleotide chain from the oligo acceptor substrate. In some embodiments, a nuclease is used to cleave or release the growing or completed oligonucleotide chain from the oligo acceptor substrate.

[0328] In further embodiments, any of the above-described processes for the conversion of one or more substrate compounds to product compound can further comprise one or more steps selected from: extraction; isolation; purification; and crystallization of product compound. As is known to those skilled in the art, acidic compounds such as oligonucleotides, NTPs, modified NTPs, and NTP-3'-O-RBGs may exist in various salt forms that can be used interchangeably in the methods described herein. All such forms are specifically envisaged for use in the methods described herein. Methods, techniques, and protocols for extracting, isolating, purifying, and/or crystallizing the product from biocatalytic reaction mixtures produced by the above disclosed processes are known to the ordinary artisan and/or accessed through routine experimentation. Additionally, illustrative methods are provided in the Examples below.

[0329] Various features and embodiments of the invention are illustrated in the following representative examples, which are intended to be illustrative, and not limiting.

EXAMPLES

[0330] The following Examples, including experiments and results achieved, are provided for illustrative purposes only and are not to be construed as limiting the present disclosure. Indeed, there are various suitable sources for many of the reagents and equipment described below. It is not intended that the present disclosure be limited to any particular source for any reagent or equipment item.

[0331] In the experimental disclosure below, the following abbreviations apply: M (molar); mM (millimolar), μM and uM (micromolar); nM (nanomolar); mol (moles); gm and g (gram); mg (milligrams); ug and μg (micrograms); L and 1 (liter); ml and mL (milliliter); cm (centimeters); mm (millimeters); μM and μιη (micrometers); sec. (seconds); min(s) (minute(s)); h(s) and hr(s) (hour(s)); U (units); MW (molecular weight); rpm (rotations per minute); psi and PSI (pounds per square inch); °C (degrees Celsius); RT and rt (room temperature); CV (coefficient of variability); CAM and cam (chloramphenicol); PMBS (polymyxin B sulfate); IPTG (isopropyl β-D-1-thiogalactopyranoside); LB (lysogeny broth); TB (terrific broth); SFP (shake flask powder); CDS (coding sequence); DNA (deoxyribonucleic acid); RNA (ribonucleic acid); nt (nucleotide; polynucleotide); aa (amino acid;

polypeptide); E. coli W3110 (commonly used laboratory E. coli strain, available from the Coli Genetic Stock Center [CGSC], New Haven, CT); HTP (high throughput); HPLC (high pressure liquid chromatography); HPLC-UV (HPLC-Ultraviolet Visible Detector); 1H NMR (proton nuclear magnetic resonance spectroscopy); FIOPC (fold improvements over positive control); Sigma and Sigma-Aldrich (Sigma-Aldrich, St. Louis, MO; Difco (Difco Laboratories, BD Diagnostic Systems, Detroit, MI); Microfluidics (Microfluidics, Westwood, MA); Life Technologies (Life Technologies, a part of Fisher Scientific, Waltham, MA); Amresco (Amresco, LLC, Solon, OH); Carbosynth (Carbosynth, Ltd., Berkshire, UK); Varian (Varian Medical Systems, Palo Alto, CA); Agilent (Agilent Technologies, Inc., Santa Clara, CA); Infors (Infors USA Inc., Annapolis Junction, MD); and Thermotron (Thermotron, Inc., Holland, MI).

Abbreviations for modified nucleotides

Abbreviations for modified nucleotides	
Alias	**Modification**
5'-FAM	5'-6-carboxyfluorescein modification
T{n}	poly-thymidine of length n
3'P	3'-phosphorylation modification
mA	2'-methoxyadenosine
mC	2'-methoxycytidine
mG	2'-methoxyguanosine
mU	2'-methoxyuridine
fA	2'-fluoroadenosine
fC	2'-fluorocytidine
fG	2'-fluoroguanosine
fU	2'-fluorouridine
rA	adenosine ribonucleotide
rC	cytidine ribonucleotide
rG	guanosine ribonucleotide
rU	uridine ribonucleotide
*	phosphorothioate linkage
5'-P	5'-phosphorylation modification
/sgna-A/	adenosine (S)-glycol nucleic acid
/sgna-C/	cytidine (S)-glycol nucleic acid
/sgna-G/	guanosine (S)-glycol nucleic acid
/sgna-T/	thymidine (S)-glycol nucleic acid
/sgna-U/	uridine (S)-glycol nucleic acid
+A	C2'-endo locked adenosine ribonucleotide
+C	C2'-endo locked cytidine ribonucleotide
+G	C2'-endo locked guanosine ribonucleotide
+T	C2'-endo locked thymidine ribonucleotide
+U	C2'-endo locked uridine ribonucleotide
(MOE)A	2'-O-(2-methoxyethyl)adenosine
(MOE)C	2'-O-(2-methoxyethyl)cytidine

Abbreviations for modified nucleotides	
Alias	**Modification**
(MOE)G	2'-O-(2-methoxyethyl)guanosine
(MOE)U	2'-O-(2-methoxyethyl)uridine
ddA	2',3'-dideoxyadenosine
ddC	2',3'-dideoxycytidine
ddG	2',3'-dideoxyguanosine
ddT	2',3'-dideoxythymidine
ddU	2',3'-dideoxyuridine

Example 1

Shake Flask Expression and Purification of TnTs and IPP

Shake Flask Expression

[0332] Evolved terminal nucleotidyl transferase (TnT) variants and wildtype inorganic pyrophosphatase (IPP) described in PCT/US2023/076667 were streaked out onto agar plates containing 30 µg/mL chloramphenicol and 0.01% glucose (v/v) for single colonies. Then, a single colony was used to inoculate a 5-mL tube containing Terrific Broth (TB) media (Teknova) and 30 µg/mL chloramphenicol, which was then shaking for 20 hours at 30°C, 250 rpm. A glycerol stock was then prepared by mixing the overnight culture with 50% glycerol at a 2:1 ratio and stored at -70°C. The glycerol stocks were thawed and 30 µL from each glycerol stock was used to inoculate 1 L shake flasks containing: 160 mL Terrific Broth (TB) media (Teknova), 30 µg/mL chloramphenicol, 0.03% (v/v) lactose and 0.075% (v/v) glucose. The shake flasks were grown at 32 °C for 18 hours at 250 rpm. Following this incubation period, the cultures were centrifuged at 4,000 rpm for 10 min at 4 °C. The culture supernatant was discarded, and the pellets were resuspended in 30 mL of 50 mM Tris-HCl, pH 7.5. The cell suspension was chilled in an ice bath and lysed using a Microfluidizer cell disruptor (Microfluidics M-110L). The crude lysate was pelleted by centrifugation (10,000 rpm for 90 min at 4°C) in an Avanti J-series centrifuge and JLA-16.250 fixed rotor (Beckman Coulter). The supernatant was then filtered through a 0.2 µm PES membrane to further clarify the lysate. The clarified lysates were then supplemented with 20 mM imidazole and 500 mM NaCl.

Purification of Enzymes from Shake Flask Lysates

[0333] Enzyme lysates were supplemented with 1/10th volume of SF elution buffer (50 mM Tris-HCl, 500 mM NaCl, 250 mM imidazole, 0.02% v/v Triton X-100 reagent) per well. Lysates were then purified using an AKTA Start purification system and a 5 mL HisTrap FF column (GE Healthcare) using the AC Step HiF setting (the run parameters are provided below). The SF wash buffer comprised 50 mM Tris-HCl, 300 mM NaCl, 20 mM imidazole, 0.02% v/v Triton X-100 reagent.

Table 3.1 Purification Parameters	
Parameter	**Volume**
Column volume	5 mL
Flow rate	8 mL/min
Pressure limit	0.3 MPa
Sample volume	35 mL
Equilibration volume	5 column volumes (CV) = 25 mL
Wash Unbound volume	20 CV = 100 mLs
Elution	Isocratic (step)
Elution volume	5 CV = 25 mLs
Fraction volume	1.5 mLs
RE-equilibration volume	5 CV = 25 mLs

[0334] Elution fractions containing protein were identified by UV absorption (A280) and pooled, then dialyzed overnight in dialysis buffer (20 mM Tris-HCl, pH 7.4, 100 mM KCl, 0.1 mM EDTA, and 50% glycerol) in a 3.5K Slide-A-Lyzer™ dialysis cassette (Thermo Fisher) for buffer exchange. Enzyme concentrations in the preparations were measured by absorption at 280 nm.

Example 2

Capillary electrophoresis (CE) analysis of oligonucleotides

Sample preparation for reaction analysis using CE:

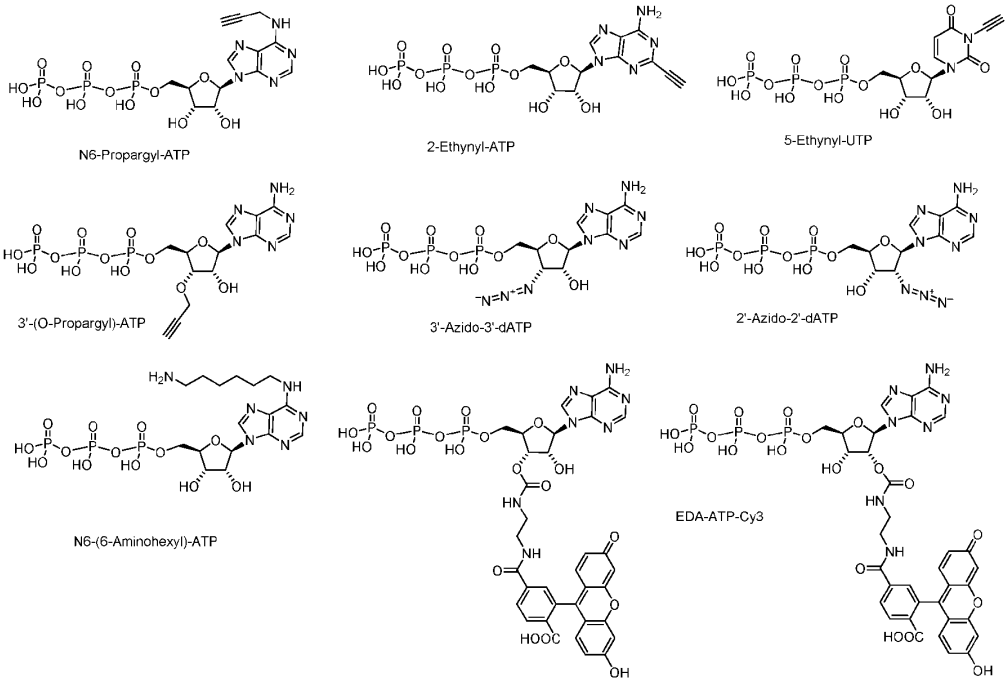
[0335] For analysis of the reaction samples, capillary electrophoresis was performed using an ABI 3500xl Genetic Analyzer (ThermoFisher). Reactions (20 µL) were quenched by the addition of 60 µL of 35 mM aqueous EDTA. Reactions (1 µL) were quenched by the addition of 99 µL of 1 mM aqueous EDTA. Quenched reactions were diluted in water to 1.25 nM oligonucleotide, and a 2-µL aliquot of this solution was transferred to a new 96-well MicroAmp Optical PCR plate or 384-well MicroAmp Optical PCR plate containing 18 µL Hi-Di™ Formamide (ThermoFisher) containing an appropriate size standard (LIZ or Alexa633). The ABI3500xl was configured with POP6 polymer, 50 cm capillaries, and a 55 °C oven temperature. Pre-run settings were 18KV for 50 sec. Injection was 10KV for 2 sec, and the run settings were 19KV for 620 sec. FAM-labeled oligo substrates and products were identified by their sizes relative to the sizing ladder.

Example 3

TnT activity with base and sugar modified nucleotide triphosphates

Activity of shake-flask purified TnT with base and sugar modified nucleotide triphosphates

[0336] TnT variants SEQ ID NO: 4, 6, and 8 were selected to evaluate a panel of base and sugar modified NTPs. The NTP panel included nucleotide triphosphates with modified base and sugar motifs including alkynyl, azido, and amino amenable to further synthetic functionalization. Activity with an NTP bearing a large fluorescent label attached to the sugar was also evaluated. NTPs were purchased from Jena Bioscience, Inc.



Scheme 3. Panel of NTPs screened with evolved TnT variants.

[0337] Reactions were performed in 96-well format 200 µL BioRad PCR plates. Reactions included 200 µM oligonucleotide, 400 µM NTP, 5 µM TnT, 1 µM SEQ ID NO: 2, 100 mM triethanolamine (pH 7.8), and 1000 µM cobalt (II) chloride. The reactions were set up as follows: (i) all reaction components, except for TnT, were pre-mixed in a single solution, and were aliquoted into each well of the 96-well plates (ii) TnT solution was then added into the wells to initiate the reaction. The reaction plate was heat-sealed with a peelable aluminum seal and incubated in a thermocycler at the indicated temperature and reaction time, then held at 4 °C until the reaction was quenched. Reactions were quenched and processed for CE analysis as described in Example 2.

Table 3.1
Reaction buffer-100 mM TEA, 1 mM $CoCl_2$, pH 7.8; IPP SEQ ID NO: 2; IPP (µM)- 1; Reaction temperature (C)- 50; Reaction Time (min)- 60 min; Reaction volume (µL)- 1; TnT concentration (µM)- 5; NTP- N6-Propargyl-rATP, 2-Ethynyl-rATP, 5-Ethynyl-rUTP, 3'-(O-Propargyl)-rATP, 3'-Azido-3'-rATP, 2'-Azido-2'-dATP, N6-(6-Aminohexyl)-rATP, EDA-rATP-Cy3; NTP conc (µM)-400; Oligonucleotide- 5'-6-FAM-T9mUmCmAmUmCmU; Oligonucleotide conc (µM)-200; Products- Extension products (+1, +2, +3 or >3 of) of 5'-6-FAM-T46mUmCmUmU with the indicated NTPs as observed by CE analysis.

[0338] Percent conversion was calculated as the percent product of the variant, defined as the sum of the area of products divided by the sum of the total peak area. The results are shown in Table 3.2-3.4.

Table 3.2

	Percent conversion to product by SEQ ID NO: 4 (See Table 3.1 for list of products detected.)		
NTP	**N+1**	**N+2**	**N+3 and >N+3**
N6-Propargyl-ATP	++	++	+
2-Ethynyl-ATP	+++	++	+
5-Ethynyl-UTP	0	+++	+
3'-(O-Propargyl)-ATP	+++	0	0
3'-Azido-3'-rATP	+++	0	0
2'-Azido-2'-dATP	++	++	++
N6-(6-Aminohexyl)-ATP	++	++	++
EDA-ATP-Cy3	+++	0	0
Levels of activity were determined by the % conversion to product and defined as follows:"+" from 1-10, "++" from 11-60, "+++" from 60-100.			

Table 3.3

	Percent conversion to product by SEQ ID NO: 6 (See Table 3.1 for list of products detected.)		
NTP	**N+1**	**N+2**	**N+3 and >N+3**
N6-Propargyl-ATP	++	++	+
2-Ethynyl-ATP	+++	++	+
5-Ethynyl-UTP	+	+++	+
3'-(O-Propargyl)-ATP	+++	0	0
3'-Azido-3'-rATP	+++	0	0
2'-Azido-2'-dATP	++	++	++
N6-(6-Aminohexyl)-ATP	++	++	++
EDA-ATP-Cy3	+++	0	0
Levels of activity were determined by the % conversion to product and defined as follows:"+" from 1-10, "++" from 11-60, "+++" from 60-100.			

Table 3.4

	Percent conversion to product by SEQ ID NO: 8 (See Table 3.1 for list of products detected)		
NTP	**N+1**	**N+2**	**N+3 and >N+3**
N6-Propargyl-ATP	++	++	+
2-Ethynyl-ATP	+++	++	+
5-Ethynyl-UTP	0	++	+++
3'-(O-Propargyl)-ATP	+++	0	0
3'-Azido-3'-rATP	+++	0	0
2'-Azido-2'-dATP	++	++	++

Table 3.4			
	Percent conversion to product by SEQ ID NO: 8 (See Table 3.1 for list of products detected)		
NTP	N+1	N+2	N+3 and >N+3
N6-(6-Aminohexyl)-ATP	++	++	++
EDA-ATP-Cy3	+++	0	0
Levels of activity were determined by the % conversion to product and defined as follows:"+" from 1-10, "++" from 11-60, "+++" from 60-100.			

Example 4

Activity improvements of evolved TnT variants relative to wildtype with a modified oligo acceptor and modified nucleotide triphosphates

[0339] TnT variants SEQ ID NO: 4 and 12 and wildtype SEQ ID NO: 10 were selected to evaluate a panel of base and sugar modified NTPs. The NTP panel included nucleotide triphosphates with modified base and sugar motifs including alkynyl, azido, and amino amenable to further synthetic functionalization. Activity with an NTP bearing a large fluorescent label attached to the sugar was also evaluated. NTPs were purchased from Jena Bioscience, Inc. See Scheme 3 for the structures of NTPs evaluated.

[0340] Reactions were performed in 96-well format 200 µL BioRad PCR plates. Reactions included 200 µM oligonucleotide, 400 µM NTP, 5 µM TnT, 1 µM SEQ ID NO: 2, 100 mM triethanolamine (pH 7.8), and 1000 µM cobalt (II) chloride. The reactions were set up as follows: (i) all reaction components, except for TnT, were pre-mixed in a single solution, and were aliquoted into each well of the 96-well plates (ii) TnT solution was then added into the wells to initiate the reaction. The reaction plate was heat-sealed with a peelable aluminum seal and incubated in a thermocycler at the indicated temperature and reaction time, then held at 4 °C until the reaction was quenched. Reactions were quenched and processed for CE analysis as described in Example 2.

Table 4.1
Reaction buffer-100 mM TEA, 1 mM $CoCl_2$, pH 7.8; IPP SEQ ID NO: 2; IPP (µM)- 1; Reaction temperature (C)- 50; Reaction Time (min)- 30 min; Reaction volume (µL)- 1; TnT concentration (µM)- 5; NTP- N6-Propargyl-rATP, 2-Ethynyl-rATP, 5-Ethynyl-rUTP, 3'-(O-Propargyl)-rATP, 3'-Azido-3'-rATP, 2'-Azido-2'-dATP, N6-(6-Aminohexyl)-rATP, EDA-rATP-Cy3; NTP conc (µM)-400; Oligonucleotide- 5'-6-FAM-T9mUmCmAmUmCmU; Oligonucleotide conc (µM)-200; Products- Sum of all extension products (+1, +2, +3, etc..) as observed by CE analysis.

[0341] Percent conversion was calculated as the percent product of the variant, defined as the sum of the area of products divided by the sum of the total peak area times one hundred. The fold

improvement over the parent (FIOP) wildtype enzyme (SEQ ID: NO 10) was determined from the ratio of the % product with the TnT variant over the % product with the wildtype enzyme. The results are shown in **Table 4.2**.

Table 4.2

	FIOP % product over the wildtype enzyme (SEQ ID NO: 10) (See Table 4.1 for list of products detected.)	
NTP	**SEQ ID NO: 12**	**SEQ ID NO: 4**
N6-Propargyl-ATP	++	++
2-Ethynyl-ATP	+++	+++
5-Ethynyl-UTP	+	+
3'-(O-Propargyl)-ATP	+	++
3'-Azido-3'-rATP	+	++
2'-Azido-2'-dATP	++	++
N6-(6-Aminohexyl)-ATP	++	++
EDA-ATP-Cy3	+	+++
Levels of activity were determined by the FIOP % product by variant TnT/ %product by wildtype defined as follows:"+" from 5-19, "++" from 20-49, "+++" from 50-85.		

[0342] While the invention has been described with reference to the specific embodiments, various changes can be made and equivalents can be substituted to adapt to a particular situation, material, composition of matter, process, process step or steps, thereby achieving benefits of the invention without departing from the scope of what is claimed.

[0343] For all purposes in the United States of America, each and every publication and patent document cited in this disclosure is incorporated herein by reference as if each such publication or document was specifically and individually indicated to be incorporated herein by reference. Citation of publications and patent documents is not intended as an indication that any such document is pertinent prior art, nor does it constitute an admission as to its contents or date.

CLAIMS

What is claimed is:

1. A method for template-independent synthesis of an oligonucleotide, the method comprising:

a) providing at least one terminal deoxynucleotidyl transferase or template-independent polymerase;

b) providing at least one oligo acceptor substrate, wherein the oligo acceptor substrate comprises a 3'-OH, or an equivalent thereof;

c) contacting the oligo acceptor substrate, the terminal deoxynucleotidyl transferase or template-independent polymerase, and a modified nucleotide triphosphates, or modified NTP-3'-O-RBG under conditions sufficient for the addition of the modified nucleotide or modified nucleotide-3'-O-RBG to the 3' end of the oligo acceptor substrate,

wherein the modified nucleotide triphosphate or modified NTP comprises a conjugate moiety, reactive group, or linker.

2. The method of Claim 1, wherein the method further comprises (d) deblocking the oligonucleotide formed in step (c) at the protected 3'-O-position of the oligonucleotide product.

3. The method of Claims 1 or 2, wherein the method further comprises e) deactivating unreacted modified nucleotide triphosphates or modified NTP-3'-O-RBGs.

4. The method of any of Claims 1-3, wherein the method further comprises an optional step (f) of removing excess nucleoside and/or excess inorganic phosphate or pyrophosphate from the reaction.

5. The method of any of Claims 1-4, wherein the method further comprises repeating steps (a)-(c) or (a)-(d) or (a)-(e) or (a)-(f) until a desired oligonucleotide sequence is obtained.

6. The method of any of Claims 1-5, wherein the method further comprises (g) cleaving or releasing the growing or completed oligonucleotide chain.

7. The method of any of Claims 1-6, wherein step (d) deblocking the oligonucleotide formed in step (c) at the protected 3'-O-position and step (e) deactivating the unreacted nucleotide triphosphates, modified nucleotide triphosphates, or NTP-3'-O-RBGs occur simultaneously.

8. The method of any of Claims 1-7, wherein the terminal deoxynucleotidyl transferase or template-independent polymerase comprises a polypeptide sequence having at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence selected from SEQ ID NO: 4, 6, 8, 10, and/or 12, or a functional fragment thereof, wherein said engineered terminal nucleotidyl transferase is capable of incorporating a modified NTP comprising a conjugate moiety, reactive group, or linker onto an oligo acceptor substrate and/or wherein said

engineered terminal nucleotidyl transferase is capable of incorporating a modified NTP onto an oligo acceptor substrate comprising a conjugate moiety, reactive group, or linker.

9. The method of any one of Claims 1-8, wherein the terminal deoxynucleotidyl transferase or template-independent polymerase comprises the engineered terminal nucleotidyl transferase of any one of Claims 12-27.

10. The method of any of Claims 1-9, wherein the conjugate moiety, reactive group, or linker comprises N6-Propargyl-rATP, 2-Ethynyl-rATP, 5-Ethynyl-rUTP, 3'-(O-Propargyl)-rATP, 3'-Azido-3'-rATP, 2'-Azido-2'-dATP, N6-(6-Aminohexyl)-rATP, EDA-rATP-Cy3, saccharide, lectin, lipid, sterol, peptide, protein, antibody, aptamer, peptide nucleic acid, vitamin, toxin, hormone, azido (N_3), amino (NH_3), bicyclo[6.1.0]nonyne (BCN), cyano (CN), dibenzocyclooctynyl, alkynyl, tetrazinyl, vinyl, or another group.

11. The method of any of Claims 1-10, wherein the terminal deoxynucleotidyl transferase or template-independent polymerase is immobilized.

12. An engineered terminal nucleotidyl transferase comprising a polypeptide sequence having at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence selected from SEQ ID NO: 4, 6, 8, 10, and/or 12, or a functional fragment thereof, wherein said engineered terminal nucleotidyl transferase is capable of incorporating a NTP comprising a conjugate moiety, reactive group, or linker onto an oligo acceptor substrate.

13. An engineered terminal nucleotidyl transferase comprising a polypeptide sequence having at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence selected from SEQ ID NO: 4, 6, 8, 10, and/or 12, or a functional fragment thereof, wherein said engineered terminal nucleotidyl transferase is capable of incorporating a NTP onto an oligo acceptor substrate comprising a conjugate moiety, reactive group, or linker.

14. An engineered terminal nucleotidyl transferase comprising a polypeptide sequence having at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence selected from SEQ ID NO: 4, 6, 8, 10, and 12, or a functional fragment thereof, and comprising at least one amino acid residue difference relative to the reference sequence.

15. The engineered terminal nucleotidyl transferase of any of Claims 12-14, wherein the polypeptide sequence comprises one or more residue differences relative to the reference sequence of SEQ ID NO: 10 at one or more of amino acid positions selected from: 12; 14; 26; 30; 34; 37; 38; 53; 79; 80; 81; 90; 92; 94; 101; 108; 121; 137; 140; 141; 142; 150; 152; 153; 155; 160; 163; 165; 174; 177; 184; 185; 189; 190; 194; 196; 201; 203; 205; 213; 219; 231; 244; 248; 258; 263; 264; 273; 284; 288; 289; 290; 293; 300; 304; 307; 313; 314; 315; 317; 318; 324; 325; 333; 336; 340; 342; 344; 352;

353; 359; 362; 379; 390; 392; 393; 394; 397; 398; 401; 402; 403; 406; 408; 410; 411; 413; 414; 415; 416; 425; 427; 428; 429; 431; 433; 434; 441; 442; 444; 446; 455; 456; 460; 461; 462; 466; 468; 470; 474; 476; 481; 484; 485; 488; 495; 499; 501; 502; 506; 515; 522; 523; and 525; and/or any combination thereof.

16. The engineered terminal nucleotidyl transferase of any of Claims 12-14, wherein the polypeptide sequence comprises one or more residue differences relative to the reference sequence of SEQ ID NO: 10 at one or more of amino acid positions selected from: 12; 14; 26; 30; 34; 37; 38; 53; 79; 80; 81; 90; 92; 94; 101; 108; 121; 137; 140; 141; 142; 150; 152; 153; 155; 160; 163; 165; 174; 177; 184; 185; 189; 190; 193; 196; 201; 203; 205; 213; 219; 231; 244; 248; 258; 263; 264; 267; 273; 274; 281; 284; 288; 289; 290; 293; 300; 301; 302; 304; 307; 313; 314; 315; 317; 318; 324; 325; 333; 336; 340; 342; 344; 347; 352; 353; 354; 359; 362; 379; 390; 391; 392; 393; 394; 395; 397; 398; 400; 401; 402; 406; 408; 410; 411; 413; 414; 415; 416; 425; 427; 428; 429; 431; 433; 434; 441; 442; 444; 446; 451; 455; 456; 459; 460; 461; 462; 466; 468; 470; 474; 476; 481; 484; 485; 488; 495; 499; 501; 502; 504; 506; 515; 522; 523; 525; and 526; and/or any combination thereof.

17. The engineered terminal nucleotidyl transferase of any of Claims 12-14, wherein the polypeptide sequence comprises one or more residue differences relative to the reference sequence of SEQ ID NO: 10 at one or more of amino acid positions selected from: 12; 14; 26; 30; 34; 37; 38; 53; 79; 80; 81; 90; 92; 94; 101; 108; 121; 137; 140; 141; 142; 150; 152; 153; 155; 160; 163; 165; 174; 177; 184; 185; 189; 190; 193; 196; 200; 201; 203; 205; 213; 219; 231; 244; 248; 258; 263; 264; 273; 274; 277; 281; 284; 288; 289; 290; 293; 300; 301; 302; 304; 307; 313; 314; 315; 317; 318; 324; 325; 333; 336; 340; 342; 344; 347; 352; 353; 354; 359; 362; 379; 380; 381; 390; 391; 392; 393; 394; 395; 397; 398; 400; 401; 402; 404; 406; 408; 410; 411; 413; 414; 415; 416; 425; 427; 428; 429; 431; 433; 434; 441; 442; 444; 446; 455; 456; 459; 460; 461; 462; 466; 468; 470; 474; 476; 477; 481; 484; 485; 488; 492; 495; 499; 501; 502; 504; 506; 515; 522; 523; 525; and 526; and/or any combination thereof.

18. The engineered terminal nucleotidyl transferase of any of Claims 12-14, wherein the polypeptide sequence comprises one or more residue differences relative to the reference sequence of SEQ ID NO: 10 at one or more of amino acid positions selected from: 12; 14; 26; 30; 34; 37; 38; 53; 79; 80; 81; 90; 92; 94; 101; 108; 121; 137; 140; 141; 142; 150; 152; 153; 155; 160; 163; 165; 174; 177; 184; 185; 189; 190; 193; 196; 200; 201; 203; 205; 208; 213; 219; 231; 244; 248; 258; 263; 264; 273; 274; 277; 281; 284; 288; 289; 290; 293; 296; 299; 300; 301; 302; 304; 307; 313; 314; 315; 317; 318; 319; 324; 325; 333; 336; 340; 342; 343; 344; 347; 352; 353; 354; 359; 362; 363; 379; 380; 381; 390; 391; 392; 393; 394; 395; 397; 398; 400; 401; 402; 404; 406; 407; 408; 410; 411; 413; 414; 415; 416; 425; 427; 428; 429; 431; 433; 434; 441; 442; 444; 446; 455; 456; 459; 460; 461; 462; 465; 466; 468; 470; 474; 476; 477; 481; 484; 485; 488; 492; 495; 499; 501; 502; 504; 506; 515; 522; 523; 525; and 526; and/or any combination thereof.

19. The engineered terminal nucleotidyl transferase of any of Claims 12-18, wherein the polypeptide sequence further comprises an N-terminal truncation of 1-156 amino acids.

20. The engineered terminal nucleotidyl transferase of any of Claims 12-19, wherein said engineered terminal nucleotidyl transferase is capable of template-independent synthesis.

21. The engineered terminal nucleotidyl transferase of any of Claims 12-20, having at least one improved property, as compared to a wild-type or reference terminal nucleotidyl transferase or template-independent polymerase.

22. The engineered terminal nucleotidyl transferase of Claim 21, wherein said improved property is selected from increased specific activity on NTP-3'-O-RBGs or natural or modified NTP substrates comprising a conjugate moiety, reactive group, or linker; increased incorporation efficiency in extension of oligo acceptor substrates comprising a conjugate moiety, reactive group, or linker; and increased activity on various oligo acceptor substrates comprising a conjugate moiety, reactive group, or linker as compared to a wild-type TnT (or TdT) or other TnTs or template-independent polymerases known to those of skill in the art.

23. The engineered terminal nucleotidyl transferase of any of Claims 12-22, wherein said engineered terminal nucleotidyl transferase comprises increased specific activity on NTP-3'-O-RBGs or natural or modified NTP substrates comprising a conjugate moiety, reactive group, or linker as compared to a wild-type or reference terminal nucleotidyl transferase or template-independent polymerase.

24. The engineered terminal nucleotidyl transferase of any of Claims 12-22, wherein said engineered terminal nucleotidyl transferase comprises increased incorporation efficiency in extension of oligo acceptor substrates comprising a conjugate moiety, reactive group, or linker as compared to a wild-type or reference terminal nucleotidyl transferase or template-independent polymerase.

25. The engineered terminal nucleotidyl transferase of any of Claims 12-22, wherein said engineered terminal nucleotidyl transferase comprises increased activity on one or more oligo acceptor substrates comprising one or more conjugate moieties, reactive groups, or linkers as compared to a wild-type or reference terminal nucleotidyl transferase or template-independent polymerase.

26. The engineered terminal nucleotide transferase of any of Claims 12, 13, and 15-25, wherein said one or more conjugate moieties, reactive groups, or linkers comprises N6-Propargyl-rATP, 2-Ethynyl-rATP, 5-Ethynyl-rUTP, 3'-(O-Propargyl)-rATP, 3'-Azido-3'-rATP, 2'-Azido-2'-dATP, N6-(6-Aminohexyl)-rATP, EDA-rATP-Cy3, saccharide, lectin, lipid, sterol, peptide, protein, antibody, aptamer, peptide nucleic acid, vitamin, toxin, hormone, azido (N_3), amino (NH_3),

bicyclo[6.1.0]nonyne (BCN), cyano (CN), dibenzocyclooctynyl, alkynyl, tetrazinyl, vinyl, or another group.

27. The engineered terminal nucleotidyl transferase of any of Claims 22-25, wherein said terminal nucleotidyl transferase comprises 2-fold, 5-fold, 20-fold, 50-fold or more the activity of a wild-type or reference TnT.

28. A polynucleotide sequence encoding at least one engineered terminal nucleotidyl transferase of any of Claims 12-27.

29. A polynucleotide sequence comprising at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to the reference sequence of SEQ ID NOs: 3, 5, 7, 9, and/or 11 or a functional fragment thereof, wherein said polynucleotide sequence encodes an engineered polypeptide comprising at least one substitution at one or more amino acid positions.

30. The polynucleotide sequence of Claim 29, wherein said polynucleotide sequence encodes at least one engineered terminal nucleotidyl transferase comprising a sequence having at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to the reference sequence of SEQ ID NOs: 4, 6, 8, 10, and 12.

31. The polynucleotide sequence of Claim 30, wherein said sequence comprises SEQ ID NOs: 3, 5, 7, or 11.

32. The polynucleotide sequence of any of Claims 29-31, wherein said polynucleotide sequence is operably linked to a control sequence.

33. The polynucleotide sequence of Claim 29 or 30, wherein said polynucleotide sequence is codon-optimized.

34. An expression vector comprising at least one polynucleotide sequence of any of Claims 28-33.

35. A host cell comprising at least one expression vector of Claim 34.

36. A method of producing an engineered terminal nucleotidyl transferase polypeptide in a host cell comprising culturing a host cell of Claim 35, under suitable culture conditions, such that at least one engineered terminal nucleotidyl transferase is produced.

37. The method of Claim 36, further comprising recovering at least one engineered terminal nucleotidyl transferase from the culture and/or host cells.

38. The method of Claim 36 or 37, further comprising the step of purifying said at least one engineered terminal nucleotidyl transferase.

39. A composition comprising at least one engineered terminal nucleotidyl transferase of any of Claims 12-27.